VOL. XLI
No. 2

PSYCHOLOGICAL REVIEW PUBLICATIONS

WHOLE NO. 185
1931

# Psychological Monographs



---

# THE INFLUENCE OF TRAINING ON CHANGES IN VARIABILITY IN ACHIEVEMENT

BY

HOMER B. REED

PROFESSOR OF PSYCHOLOGY, FORT HAYS KANSAS STATE COLLEGE,
HAYS, KANS.

---

PUBLISHED FOR THE AMERICAN PSYCHOLOGICAL ASSOCIATION BY

PSYCHOLOGICAL REVIEW COMPANY

PRINCETON, N. J.
AND ALBANY, N. Y.

TABLE OF CONTENTS

# CHAPTER I

# PREVIOUS EXPERIMENTS

## SECTION 1

### THE PROBLEM, ITS IMPORTANCE, AND POSSIBLE METHODS OF SOLUTION

Does equal training make a group of individuals more alike or more different in their achievement? More exactly, does it increase or decrease the variability or individual differences of a group of individuals? The solution to this problem is tied up not only with the old question of the relative importance of heredity and environment in the achievement of individuals, but also with questions of public policy in education. If individuals under the same training become more alike in their achievements it is said that this result is due to environment. On the other hand, if they become more different, it is said that this result is due to heredity. If our great system of public education will have the result of increasing the social, political and economic inequalities between individuals, then it fails in one of its major purposes, and many of our present educational practices accomplish just the opposite for which they are intended; but if it decreases the inequalities between individuals, then our present gospel of equality of opportunity is well grounded, for the theory underlying the latter doctrine is that inequalities between individuals will not be great if all individuals have an equal chance.

If we try to reach a solution by theoretical reasoning, we can follow either of two lines. We may assume that achievement along a given line, for example, earning money, is practically unlimited and that great ability means not only great wealth but also the ability to earn it at a rapid rate. Hence the difference between the wealth of such an individual and of one of little ability will increase during the earning period of the two, and at

the end of this period, the two individuals will be much farther apart than at the beginning. Many individual cases can be found where this theory is true to fact. On the other hand, we can compare achievement along some lines to making a journey from A to B. Two individuals making this journey travel at different rates, and the rate of the same individual varies according to conditions while on the journey but the rate of each one decreases as he approaches the limit B. At certain times the distance between the two individuals increases but these are the exceptions. In the end the two will be together and the nearer they approach the end the closer together they will be. The greater the distance from the starting point the less the distance to the end—and if we think of the untraveled distance as the gain—the smaller the gain will be both absolutely and relatively. To the writer this illustration seems to be typical of any skill that is learned by an individual. In the performance of any skill certain movements are required, and the number of these movements which can be made in a unit of time is limited, the limit being presumably set by the capacity of the nervous system to transmit impulses.

If we give the same training to a group of individuals in a certain skill, it seems reasonable to suppose that the individuals will become more alike as a result of this training. In this group let us assume that some have already learned the skill fairly well while others have not learned it at all. In view of the fact that the rate of improvement decreases as the physiological limit is approached, it would seem that the differences in achievement between such individuals would be reduced as a result of all of them receiving an equal period of systematic training. There is, however, the possibility that the physiological limits of different individuals are far apart and that the variability in the achievement of a group of individuals is much less when they are far from the limits than when they are near them.

If we reach opposite conclusions on the solution of our problem when we reason theoretically, our first thought of a method of reaching the true solution is experimentation and measurement. But an examination of the experimental literature discloses the fact that the experimental method also yields opposite conclusions.

It all depends upon what we accept as the measure of individual differences or of variability in a group. In fact we may first formulate our conclusions and then make an experiment and select a measure which proves them. It should be possible, however, to agree on valid measures of variability, and if we do, agreement on conclusions will not be difficult.

In the hope of reaching such an agreement we shall, first, review the experimental literature on this problem, emphasizing the contradictory conclusions reached by different investigators and showing the dependence of each conclusion upon the kind of measures used; second, criticize the various measures that have been used to solve our problem and develop valid measures of the effects of practice on changes in variability in achievement; third, apply these measures to the data of other investigations and show that consistent conclusions are not difficult to reach when this is done; and fourth, report two new investigations which we hope will lead to an acceptable solution.

## Section 2

### Review of Previous Experiments

Thorndike has made a number of experiments in which the influence of practice on individual differences was noticed. In each of them he draws the general conclusion that practice increases individual differences, although he notices an opposite effect in one of them. In 1908 he gave twenty-eight college students practice in mentally multiplying twenty-six three-place by three-place numbers at the rate of five or six examples a day. He scored the results by the time per example plus one-tenth of this time for each digit wrong in the answer. He calculated the ratios of the slower to the quicker individuals before and after practice. He found that the ratios of five of these pairs—28/1, 27/2, 26/3, 25/4, 24/5—showed decreases while three of the pairs—23/6, 22/7, 21/8—showed increases. His conclusion was: "In this experiment the larger individual differences

increase with equal training, showing a positive correlation of high initial ability with ability to profit by trainng. . . .

"It seems extremely probable from many facts of dynamic psychology that the man who has the capacity to improve to a given small degree more quickly than another, should also improve more quickly to the next degree and should also by and by be capable of improving to a given higher degree if given the maximum of efficient training."

In 1910, Thorndike made an experiment on practice in addition. Nineteen college students added daily for seven days forty-eight columns of ten digits each. The performance was scored by time plus a penalty for errors. As a measure of variability Thorndike calculated the correlation between initial performance and per cent of improvement and also the ratio of lowest to highest, next lowest to next highest, etc., at the beginning and end of practice. The correlation was found to be —¼ and the ratios, except that between the first and nineteenth individuals, showed decreases. In this case Thorndike concluded that the differences were due largely to environment and that equal practice did reduce a little the relative differences within the group.

In 1913 Donovan and Thorndike made a similar experiment on twenty-nine fourth-grade boys, the practice being divided into thirty two-minute periods. In this case the individuals were arranged according to initial scores from lowest to highest by groups of four, and the average gross gain was calculated for each group. The results showed a tendency for the initially lower individuals to make the smaller gains. For example, the first four gained 5.5 as against 6.8 by the seventh four. The author concluded: "If we compare the improvement of the eight boys who showed the least ability at the start with the seven who showed the most ability at the start we find that the latter made equal or greater gross gains." "These results . . . make it improbable that a very large fraction of the differences found among the school children can be justly attributed to differences in amount of training."

In 1914 Hahn and Thorndike, using the same sort of practice sheets, made an experiment in distributed practice on one hun-

dred ninety-two pupils—forty-two in the seventh grade, fifty-four in the sixth grade, fifty-three in the fifth grade, and forty-three in the fourth grade. The total amount of practice including the tests amounted to one hundred twenty minutes. When the pupils were arranged in groups according to their initial scores it was found that up to a certain point the gross gains increased with the size of the initial score. The authors concluded: "Those whose original capacity or circumstances of training are ahead at the start maintain and increase their lead. . . . Nature, not nurture, seems to be the chief cause of the differences in ability to add found in children of the same grade. Equalizing opportunity does not seem to equalize achievement."

In 1915 Thorndike reports an experiment in which one hundred sixty-three students were given practice in multiplication by substitution, writing the products of eleven by eleven up to nineteen by nineteen by the aid of a key. Every pupil did twelve sheets of eighty products each. The scores were in terms of time per sheet plus a penalty for errors. These were later converted into amount of work per unit of time. Thorndike grouped the individuals according to initial ability from highest to lowest and found that the gross gains were somewhat proportional to the initial scores. For example, Group A, which had an initial score of 15.4, gained 8.8, while Group G, which had an initial score of 6.1, made a gain of 7.5. Thorndike concludes that the status which an individual has obtained in a given amount of practice is prophetic of the amount he will obtain by additional practice. "In general the gross improvement in product produced per unit of time is greatest for those of initial highest ability. The effect of equalizing opportunity is to leave the gross variability between individuals unimpaired, or even to increase it."

In 1916 Thorndike made an experiment in which sixty-four educated adults practiced writing the products of 11 x 11 to 19 x 19 with the aid of a key. Four different sheets of eighty pairs each, were used in rotation, each one twelve times, making forty-eight sheets done or three thousand eight hundred forty entries. When the individuals were classified into groups according to their initial scores, it was found that the eight initially

highest gained 16.2 products per minute as against a gain of 4.7 by the five initially lowest.

In another experiment reported in the same paper, Thorndike had eleven individuals practice five days a week for two weeks canceling 2's, canceling 3's, adding columns, multiplying mentally three-place by three-place numbers, and typewriting. He states that " except in case of typewriting the gain in product produced per unit of time is a little greater for those of initially high ability. In regard to the effect of practice on individual differences he concludes: " The effect of equalizing opportunity is thus to increase individual differences. This result now found with many different functions, furnishes perhaps the strongest argument that differences in achievement are largely due to differences in original capacity."

During the period from 1908 to 1916 M. T. Whitley, F. L. Wells, J. C. Chapman and W. A. McCall, working under the inspiration of Thorndike, made contributions to this problem.

Whitley begins her work on the effect of practice on individual differences by giving an illuminating discussion of different measures of improvement, such as gross gain in amount; gross gain in time; expressing all scores in percentages of the first, when it is based on amount, and the same when it is based on time. She shows that none of these yields the same results but she does not evaluate them except to say that improvement is better measured by gains in amount either in time or work. She recommends four measures for measuring the effect of practice on individual differences but does not evaluate them. They are the average deviation, the average deviation divided by the average, the average deviation divided by the square root of the average, and the ratio of best to worst at the beginning and at the end of practice. In her experiment nine women college students took from 16 to 20 periods of practice in tracing a maze, cancellation, mental multiplication, sorting counters, and discriminating weights, all of which except the weight test, were scored by time per unit of work. Strange to say, she does not measure the effects of practice on individual differences by any of the measures that she recommended for the purpose, but bases her own judgments

on the similarities in the learning curves of different individuals and on the negative correlations that she found between initial performance and per cent of gain according to which she says, practice reduces individual differences.

Wells, in 1912, had ten adults practice on addition problems five minutes a day for thirty days, and on a number checking test daily for thirty days. Curves were plotted showing the absolute gains in amount of work per unit of time. These give a fan-shaped appearance, indicating divergence. Wells, however, acknowledged that if the curves were plotted on a basis of time per unit of work, they would show convergence. He also calculated the relative gains of the individuals and found a wide disagreement in the ranks obtained on this basis from those based on absolute gains. In spite of the disagreement between these three measures, Wells bases his conclusions on the absolute gains and says that a high initial score may carry with it more prospect of improvement than a low one and that original nature is largely responsible for this fact.

Chapman gave to twenty-two college students ten practice periods in color naming, canceling opposites, addition, and mental multiplication. He calculated the correlations between the initial and final performances, and between initial score and gross gain. He found the correlations between initial and final scores to range from .59 to .96, and between initial score and gross gain to range from .19 to .41, the highest positive correlations in both groups being in addition and multiplication. He concluded that in a complex function a high initial score is an index to a corresponding amount of improvement while in a simple function, it has little relation to improvement. This fact "affords support, based on actual measurements, to the old truism, 'to him that hath shall be given,' upon which the selective process of education is based."

McCall gave from fourteen to eighteen periods of practice to eighty-eight sixth grade pupils in such tests as cancellation, addition, copying addresses, and handwriting. The influence of practice on individual differences was not one of his problems, but

since he gives the results from his tests, this effect can easily be calculated according to whatever measure we may choose.

We may now summarize the results of Thorndike and his co-workers upon this problem. The conclusion of these experiments is that equalizing opportunities increases individual differences. The principal measure upon which this conclusion is based is gross or absolute gain, but it is also supported by ratios of lowest to highest or of highest to lowest at the beginning and end of practice, by a positive correlation between initial and final performance, and to some extent, by a positive correlation between initial performance and gross gain. It is contradicted by negative correlations between initial performance and per cent of improvement as found by Thorndike in his experiment of 1910 and by Whitley in her experiments of 1911, by the relative gains in improvement as found by Wells, and by absolute gains in time per unit of work as found by Wells. The validity of the conclusions is thus seen to depend on the validity of the measures, which we shall examine in a moment.

Since 1915 contributions to our problem have been made by G. C. Myers, L. L. Thurstone, F. A. C. Perrin, D. Starch, V. A. C. Henmon, Georgina S. Gates, Henrietta Race, H. B. Reed, M. C. Barlow, J. Peterson, Margaret Kincaid, and A. W. Hurd—all of which we now wish to review.

Myers in 1918 made an experiment in which twenty-seven college students sorted cards for twenty-six seconds, twenty successive times, during a period of fifty minutes. This was repeated ten days later, and again after one day, and finally after a period of three and one-half months. Fourteen other subjects fitted thirty-six cubes into a box for twenty-three successive school days. As a measure of variability, Myers used the average deviation divided by the average. The resulting coefficients showed practically no change from the beginning to the end of practice and Myers concluded, "Practice does not make individuals more or less alike."

Thurstone in 1918 analyzed the learning data obtained from one hundred sixty-five drafted men who had practiced telegraphy for seventy-two hours. He plotted the curves for the median,

first quartile, and third quartile scores on the bases of amount of work per unit of time and found them to be straight lines, except that the quartile curves diverged. He concluded: "Learners separate from each other more and more as they progress." In criticizing the experiment Peterson pointed out that judgments about the effects of practice on variability cannot be made on scores of absolute units such as amount of work per unit of time, or amount of time per unit of work. Like Wells, in his work on practice in association, he pointed out that curves based on time per unit of work converge and indicate just the opposite effect of the curves plotted by Thurstone. After calculating the standard deviation and dividing this by the average he finds that the resulting coefficients show a decrease in variability from the twentieth to the seventy-second hour of practice.

Perrin in 1918 reports an experiment in which twenty-one subjects practiced solving analogies at the rate of twenty-five at a sitting, once a week for ten weeks. The score was time per analogy. They also practiced mirror-reading typed material at the rate of eighteen pages a sitting, once a week for seven weeks. The score was time per unit, a unit consisting of one-sixth of a page. He plotted the curves for the five best and the five poorest subjects and found that they showed a decided convergence from which he concluded that "the superior subjects as contrasted with the inferior, disclosed less capacity for improvement and less tendency toward variability."

Chapman in 1919 reports a practice experiment in typewriting in which he gives the scores of a weekly five-minute test of twenty high school pupils who took a regular course in typewriting. The scores began with the twentieth hour and ended with the one hundred eightieth hour. In this paper Chapman was not interested in the effect of practice on individual differences but in learning curves. However, we have used his results for this purpose and have given the calculations in the summary table of previous investigations.

Starch in his Educational Psychology, 1919, discussed the influence of equal practice on individual differences. After giving the results of some of Thorndike's experiments he reports two of

his own. In the first, eight students practiced multiplying mentally fifty three-place numbers by a one-place number each day for fourteen days. He calculated the gross and relative gains for the three best and the three poorest persons and found that the best gained the most both relatively and absolutely. In the second, he gives the results of one hundred twenty minutes' practice in substitution by twenty students. He finds that the initially five highest gain more absolutely than the five lowest. He does not give the relative gains for this experiment but calculation shows that they are in favor of the initially lowest. He concludes as follows: "All experimental results point in the direction that practice does not equalize abilities, in fact equal practice tends to increase differences in achievement and skill rather than to decrease them. The more gifted individuals profit more, both relatively and absolutely, than the less gifted. This experimental fact is one of the most profound bits of evidence regarding the whole problem of heredity and environment. The talented men not only start with greater initial capacities but seem also to be capable of more intense application and more zealous desire to improve. 'To him that hath shall be given' is psychologically true in the sphere of intellectual training as well as in the sphere of morality and religion."

Henmon in 1920 reports on some individual differences in arithmetic. Under his supervision, the Courtis arithmetic tests were given each month throughout the school year. After arranging the pupils in tertiles or quartiles from highest to lowest according to gross gains made during the year he finds that the groups that made the highest gross gains had a higher initial score than the group that made the lowest gross gain. He concludes: "An inspection of the tables shows that any rational method of interpretation, the scriptural quotation 'to him that hath shall be given' holds. Those with the highest initial scores have the highest final scores and gain most both absolutely and relatively." In our paper published in 1924 we examined this experiment somewhat in detail and refer the reader to that source for a different interpretation of the data.

In 1922 Henrietta Race reports some interesting investigations

which have a bearing on our problem. She made experiments on ninety-five college students, forty-three grade pupils of superior intelligence, and one hundred seventy-two grade pupils of average intelligence. The college students had practice exercises in addition, multiplying by substitution, and cancellation. Group A practiced eight minutes a day for twenty days, Group B twenty minutes twice a week for eighteen weeks, and Group C twenty minutes an hour, eight hours a day for three days—each group having a total of one hundred sixty minutes. The grade pupils had the following practice exercises: addition, eight minutes a day for ten days; multiplication by substitution, five minutes a day for twenty days; language completion, ten minutes a day for ten days; reasoning processes in United States History, forty minutes a day for twenty days; and Thorndike Intelligence Tests, thirty minutes a day for fifteen days. The relation of improvement to initial score was measured by comparing the average gross gains of the highest and lowest fourths. In case of the college students, the highest fourth made the greater gains in addition and multiplication while the lowest fourth made the greater gain in cancellation in one test and the lesser gain in the other. In case of the grade pupils the highest fourth made the greater gains in addition and in multiplication while the lowest fourth made the greater gains in language, history, and intelligence tests. The author concludes: "On the whole then our findings are in fairly close accord with those of previous students, and all indicate that the ability possessed by any person at any time is in a large measure a product of what native capacity he has and a prophecy of what further improvement he will make from a given amount of practice."

In the same year that Miss Race published her thesis, Georgina S. Gates published an intensive study which contained results and conclusions which were for the most part contrary to those hitherto published by other investigators. She had twenty-three women college students practice from twenty-two to twenty-nine periods on color naming, tapping, addition, mental addition, and word-building. One distinctive feature of her study was the number of ways in which she measured improvement. Using the

difference between the medians of the first and last three trials as a basis, she calculated improvement by the gross reduction in amount of time per unit of work, the gross increase in amount of work per unit of time, the percentage of gains from the first trials, and the increase in rank from the first to the last trials. She also calculated improvement by each of these measures using the difference between the first and best trials as a base. Another contribution of her study was the inclusion of the computation made on the practice data obtained by Hollingworth in his study of the effect of caffeine. These are very valuable for the reason that Hollingworth had his subjects practice until the limits of improvement were reached. The results used from this study were obtained from the following tests: Color naming, opposites, addition (calculation), tapping, weight discrimination, crossing (canceling), and the three hole test. Each of the measures of improvement was correlated with the initial performance. The following results were obtained for the time and percentage measures by the formula, $= \frac{6 \Sigma D^2}{N(N^2-1)}$

INVESTIGATION A (Gates)

| Test | Gross time | Percentage |
|---|---|---|
| Color naming | —.31 | —.24 |
| Tapping | —.15 | —.15 |
| Adding | —.11 | .16 |
| Multiplication | —.84 | —.17 |
| Word building | —.26 | —.11 |

INVESTIGATION B (Hollingworth)

| Test | Gross time | Percentage |
|---|---|---|
| Color naming | —.73 | —.51 |
| Opposites | —.98 | —.83 |
| Calculations | —.92 | —.69 |
| Tapping | —.03 | —.03 |
| Discrimination | —.53 | —.53 |
| Cross Test | —.02 | +.20 |
| Three Hole | —.89 | —.66 |

Out of sixty-four such correlations all but eight were negative. Gates also gives the average gross gains in terms of amount of work per unit of time in relation to initial score, grouped from highest to lowest. They are as follows:

| Indiv. | C.N. | Tap. | Add. | Mult. | W. Bldg. |
|---|---|---|---|---|---|
| 1–6..... | 11.63 | 31.3 | 22.2 | 101.8 | 4.65 |
| 7–12..... | 12.16 | 64.5 | 21.3 | 403.8 | 7.00 |
| 14–18..... | 14.63 | 33.8 | 18.5 | 623.0 | 8.10 |
| 19–23..... | 36.69 | 103.6 | 27.8 | 738.6 | 6.50 |

"The correlations and averages, then," Mrs. Gates says, "seem to show a general negative relationship between initial ability and improvement through practice. This is true whether we measure the gain over long or short intervals of practice, whether in general we use the time, amount, per cent, or rank method, whether the function studied is the complex multiplying or the simple color naming." . . . "A high initial score should make one predict relatively little rather than relatively great future gain." . . . "Our results may be used to show that individual differences decrease through practice, though various difficulties encountered in measuring improvement and the high correlation observed between initial and final position, may indicate that the true conclusion is the opposite of the one that the statistical procedure seems to demonstrate."

Reed in 1924 pointed out that the conclusion of leading psychologists that practice increases individual differences was false because of the measures on which it was based. He said that this conclusion was justified when gains are measured by absolute increases in amount of work per unit of time but that an entirely opposite conclusion is reached when gains are measured by time per unit of work, and that because of this contradiction, neither measure is valid. The difficulty with either measure is that psychologists have been handling them like whole members, but in reality they are fractions such as $\frac{G}{T}$ or $\frac{T}{G}$ where T and G represent time and gain respectively, and that fractions cannot be compared by dealing only with the numerators. He also said that correlation between initial and final performance could not be used as a measure of changes in variability, for it was entirely possible for a bright pupil to gain relatively less than a dull one without changing his rank just as a man with a million dollars earning 3 per cent is still richer at the end of a year than a man

with a thousand dollars earning 100 per cent. As a way out of the dilemma, he based the conclusion from his experimental results upon the following measures: the ratio of the average of the four highest to the average of the four lowest at the beginning of practice compared with the ratios of the averages of these same individuals at the end of practice, the ratio of the third quartile, Q3, to the first quartile, Q1, at the beginning of practice compared with this ratio at the end of practice; and correlation between initial performance and per cent of improvement. When the first measure was applied to the results of some of the experiments of Thorndike, Starch and Henmon, it indicated a general tendency for practice to reduce individual differences. When the three measures were applied to results obtained from over a hundred students in practice experiments upon mirror-drawing, saying alphabet backwards, substitution and addition, they justified the conclusion that training reduces individual differences. Chapman criticized the ratio method used in this experiment on the ground that scores were made on an arbitrary scale with an arbitrary zero point. "The individual who improves from the arbitrary zero to the first point of a scale in a small period of practice, has by the ratio method improved infinitely. This is the *reductio ad absurdum* of the procedure." In reply it may be stated that the ratio method, just like every other method, would increase in validity if used on a scale having an absolute zero, but when applied to scores from an arbitrary scale, it is just as valid as other measures of this sort. It is also true that the method cannot be used on zero scores but that does not invalidate its use with scores that are greater than zero.

Barlow in 1924, working under the direction of J. Peterson, wrote a master's thesis in which he reported the results of three practice experiments made on twenty twelve-year-old children. They consisted of twelve practices with digit-symbol test, six practices in card-sorting, and twelve to twenty practices with mental multiplication. As a measure of variability he used the coefficient of variability obtained by dividing the standard deviation by the average. According to this measure the card-sorting tests showed a decrease in variability while the mental multipli-

cation and substitution tests showed an increase. He concluded that the more complex the trait learned, the stronger the tendency for the individuals to become more unlike with equal practice, and the simpler it is, the stronger the tendency for them to become more alike.

In 1928 Peterson and Barlow published the coefficients of variability for the results of most of the experiments on the influence of practice on individual differences that were published before that date. They report forty pairs of coefficients from as many tests made by previous investigators. Of these, twenty-four show a reduction in variability, five show no definite change, and eleven show an increase. They also report some new experiments of their own. One of these was an experiment in mental multiplication by ninety-six school children having a median age of 175.9 months. Group A, forty-two pupils, worked thirteen units of one hundred twenty-six problems each and changed its variability, 90P–10P/md, from .28 to .19. Group B, twenty-four pupils, completed ten to twelve units and changed its variability from .20 to .21. Group C, thirty pupils, completed seven units and changed its variability from .23 to .27. Group D, which consisted of all the pupils treated as one group, changed its variability from .37 the first unit to .32 in the seventh unit. This change may be taken as the general effect of this experiment.

In a second experiment the variability coefficients are given of five standardized tests of achievement in school subjects, namely: Ayers Spelling (twenty words), Nassau County Composition, Woody-McCall Mixed Fundamentals, Monroe Arithmetic Reasoning Accuracy, and Monroe Arithmetic Reasoning Principles. These were given to one hundred ten pupils in Grades V, VI, VII and VIII in the Peabody Demonstration School in January. Three of the tests, Composition, Mixed Fundamentals, and Arithmetic Reasoning Principles show a decrease in variability from Grade V to Grade VIII, while two, Arithmetic Reasoning Accuracy and Spelling show no change. In the opinion of the writer these results show a tendency but not much reliance can be placed upon them because of changes in population from grade to grade. However, it is not clear from the description whether

these tests were given to the same children year after year or to four grades in one year.

In a third experiment, Peterson reports the changes in variability that occur in weekly objective tests in a psychology course of twelve weeks in each of four sections which had a total of 224 students. All sections underwent marked reductions in variability: Section A from .29 to .13, B from .35 to .11, C from .23 to .11, and D from .33 to .19. Peterson draws no definite conclusions but recommends that future experiments take measurements at various stages of practice and that they calculate correlations between the numerators and the denominators of the variability ratio.

In 1929 Hurd published a paper on "Re-organization in Physics" which contains data relating to our problem. He outlines a unit course in physics in which each unit is preceded and ended by an objective test. He gives the results from Unit I which was on the hydrometer, taught in a period of three weeks to classes having a total population of two hundred twenty-three pupils. The unit covered the various kinds of hydrometers, relation between English and metric units of measurement, relation between density and specific gravity, experimental determination of the densities and specific gravities of some solids and liquids, and other problems. We were interested in finding out whether the pupils were more alike in their knowledge of these topics after receiving instruction about them than before. The initial and final tests show averages of 12.21 and 33.84 respectively and corresponding S. D.'s of 8.35 and 11.95. The coefficients of variability are .68 and .32. We attach much importance to these results as they are indicative of what we may expect of the effect of equal training in the schools on the variability of the achievement of the pupils.

Because we wish to make Margaret Kincaid's paper the basis of some discussion we have reserved its review until this point. In 1925 she sought to solve the problem of the influence of practice on individual differences by applying all possible measures to most of the data on the problems that had been published up to that time and also to the results of two experiments of her

own on Braille writing and dart throwing. From these she selected eight measures upon which she based her conclusions. They were as follows: the standard deviation in the initial performance as compared with the same in the final performance, the ratio of the worst to the best in the initial performance as compared with the same in the final performance, the ratio of the next worst to the next best in the initial performance as compared with the same in the final performance, the gross gain made by the highest 25 per cent as compared with the same for the lowest 25 per cent, the correlation between initial ability and gross gain, and the correlation between ability and per cent of gain. Since it is a matter of interest to see how many of these favor increase or decrease in variability as a result of practice, Table I gives a summary of the totals.

TABLE I

*Number of measures which favor decrease or increase in variability as a result of practice in 24 experiments. From Kincaid, 1925*

| Measures | Decrease | Increase |
|---|---|---|
| S.D. | 5 | 19 |
| S.D.÷Average | 16 | 8 |
| Ratio, worst to best | 14 | 10 |
| Ratio, next worst to next best | 14 | 10 |
| Gross gain | 13 | 11 |
| Per cent in gain | 23 | 1 |
| Correlation between initial ability and gross gain | 12 | 12 |
| Correlation between initial ability and per cent of gain | 22 | 2 |
| Totals | 119 | 73 |

The conclusion drawn by Kincaid is that there is on the whole a preponderance of evidence among these cases for the judgment that differences generally decrease with practice. It is unfortunate that she did not subject the various measures to a critical analysis. To apply as many measures as possible and then vote with the majority is hardly a scientific procedure, although the conclusion is undoubtedly correct. Of the above measures, the standard deviation and gross gain clearly favor the proposition that practice increases individual differences while the coefficient of variability (S. D.÷average), per cent of gain, and correlation between initial ability and per cent of gain clearly favor just the opposite. On the other hand, no conclusion can be drawn from

the ratio of worst to best, ratio of next worst to next best, and correlation between initial ability and gross gain. These results show the necessity of analyzing the validity of the measures used for the purposes of solving this problem, which is our next task.

## Section 3

### Criticism of Previous Experiments

The experiments which we have reviewed may be criticized on two general grounds, namely: experimental conditions and measurement of results. In regard to the former, the ideal is that the practice should be the same for all members of a group; and in regard to the latter, the ideal is that the measures used should measure validly and accurately what is intended to be measured, namely, changes in variability. By sameness in practice we mean that the practice should be the same in kind of material, the same in difficulty, and the same in amount. In the above experiments, there were only approximations to these conditions. For example, practice in addition or mental multiplication meant only that the individuals added or multiplied mentally but it did not mean that all the individuals added exactly the same problems or multiplied mentally exactly the same numbers. As a matter of fact addition problems of five three-place numbers or multiplication problems of four-place by four-place numbers differ widely from each other in difficulty and in the specific operations used. Even if the practice was the same in name for the members of a group, there was usually no attempt at sameness in difficulty and in specific operations. Sameness in amount of practice is also difficult to secure. Amount was sometimes kept uniform in quantity of material and sometimes uniform in quantity of time, and often there were only approximations to each of these conditions. If the quantity of material is kept the same, it is certain that the individuals will differ widely in the amount of time required to do it, while if the time is kept constant, it is certain that the amounts done by individuals will differ widely

from one to another. We do not know which of these is the correct meaning of sameness in amount of practice, but it is unlikely that the two conditions would yield the same results. Our purpose in calling attention to these conditions is not to find fault but to explain contradictions in the results and conclusions and to point out the necessity of further researches in which these conditions are controlled. We believe, however, that the contradictions in the conclusions of previous experimenters are due more to the methods of measurement that were used than to the experimental conditions.

First of all we may rule out the standard deviation as a measure of variability for the purposes of this problem. No judgment can be made on the variability of a group on two or more tests by a simple comparison of the S. D.'s alone. An S. D. has meaning only in relation to the average from which it is computed, and states the range from the average within which two-thirds of the cases fall, but it has no meaning in relation to another S. D., when the averages are disregarded. Its size is dependent on the sizes of the average and of the cases from which it is computed. An increase in the size of the S. D., from the beginning to the end of practice does not mean that the variability of the group has increased, because the average of the group may have increased relatively more than the S. D., and because the units at one stage of practice are not directly comparable with the units of another stage of practice. Kincaid's experiment in Braille writing furnishes a good example of this. At the beginning of practice the S. D. of the performance was 29 and at the end it was 92, which is more than three times larger, but this does not mean that the variability of the group has increased, for in the mean time the average of the group increased from 102 to 543. Now if we divide the S. D.'s by their averages we can form a judgment about the variability of the group at the two stages of practice. In this case it actually decreased from .29 to .17.

The correlation between initial performance and gross gain is not a reliable measure of the variability of a group from one stage of practice to another. A pupil with a high initial score may make a larger absolute gain than a pupil with a low initial

score but his relative gain may be much smaller, and if so, there will be a positive correlation between initial performance and gross gain and also a positive correlation between initial and final performance but a negative correlation between initial performance and relative gain. If two individuals are far apart at the beginning of practice and the high one makes a smaller relative gain than the low one, the two will eventually come together. If their rates of gain remain constant, the low one will overtake the high one. The coming together at a certain stage of practice of individuals, who are far apart at the beginning, means that their variability has been reduced. Negative correlations between initial performance and gross gain or negative correlations between initial performance and relative gain would be indications of reduced variability but a positive correlation between initial performance and gross gain or a positive correlation between initial and final performance may mean either an increase or a decrease in variability. However, a positive correlation between initial performance and relative gain would be an indication of an increase in variability. Kincaid's results from Braille writing again furnish some examples. The correlation between the first and last performance was .50 and that between initial performance and gross gain was .10, but the correlation between initial performance and relative gain was —.89. We have already noticed that the coefficient of variability was reduced from .29 to .17. Here we see that positive correlations between initial and final performance and a positive correlation between initial performance and gross gain are consistent with reduced variability. In her work on dart-throwing Kincaid found a positive correlation between the first and last performance of .74 but a negative correlation of —.46 between initial performance and relative gain. In this case the coefficient of variability decreased from .38 to .31. In the two experiments on dart throwing and Braille writing we see that positive correlations between initial and final performance tell us nothing about the changes in variability. In the one case a positive correlation between initial performance and gross gain goes with reduced variability, while in the other a negative correlation between these two factors goes

with reduced variability. In both cases, negative correlations between initial performance and relative gain are consistent with decreases in the coefficient of variability.

From what has already been said about gross gains it follows that they are not reliable measures of changes in variability from one stage of practice to another. An individual with a high initial score often gains more absolutely during a given period than one with a low initial score but less relatively. If so, the variability between the two is reduced. As was stated before, the low one with the higher rate of gain will eventually meet the high one. This possibility is also shown in Kincaid's results from Braille writing. The gross gain of the highest 25 per cent was 467 and that of the lowest 25 per cent was 435, but the percentage of gain of the highest 25 per cent was 334 as against one of 621 by the lowest 25 per cent of the group. Hence it is clear that these two sections were closer together at the end of practice than at the beginning and this is indicated by the reduced coefficient of variability and a negative correlation of —.89 between initial performance and relative gain.

The ratios of best to worst or of next best to next worst, etc., are not reliable measures of changes in variability from one stage of practice to another. This measure introduced by Thorndike and used as the basis of the conclusions in several of his experiments is subject to the following errors: It compares only the extremes of the group; and what is true of the extremes of the group, may not be true of the group as a whole. But the most serious error is that it rarely happens that worst and best, or next worst and next best, at the beginning of practice continue to occupy the same positions at the end of practice. Consequently the ratio of worst to best at the end of practice usually represents a comparison of different individuals from those at the beginning of practice and tells us nothing about the convergence or divergence of the initially lowest and the initially highest individuals. If, however, we calculate the ratio of the worst to the best or *vice versa* at the beginning of practice and then calculate the ratio of these same individuals at the end of practice, we shall at least find out whether these particular individuals have a tendency to

converge or diverge during the course of practice. If this is done, we shall get valuable information, for it may very well be that those who are at the extremes of the distribution at the beginning of practice converge or diverge a great deal more during the course of practice than those near the average. Even when used in this way the ratio of worst to best has the limitation of not measuring the variability of the group as a whole. However, the ratios of worst to best, next worst to next best, etc., at the beginning and the end of practice, when taken regardless of the same individuals, are almost meaningless.

The percentage of gain by the highest 25 per cent of the group as compared with that of the lowest 25 per cent of the group is subject to some of the same criticisms as the ratio of worst to best. Not all the individuals in these groups at the end of practice will be the same as those at the beginning of practice. However, many of the shifts in rank that occur do not change groups or divisions. Hence this measure is far superior to the ratio of worst to best; for in the latter, as has been stated, the same individuals rarely occupy these positions at the beginning and the end of practice. If the percentage of gain by the highest 25 per cent is smaller than that of the lowest 25 per cent it is almost sure to be an index of reduced variability, and the opposite, if it is larger. We may, therefore, accept it as one measure of change in variability but it would be a much better measure if, instead of taking the highest and lowest 25 per cent at the end of practice, we calculated the average percentage of gain made by the individuals who constituted these groups at the beginning of practice. Such percentages would tell us whether the highest and lowest fourths converge or diverge during the course of practice. What is true of these sections, however, may not be true of the individuals in the middle half of the group but it is likely to be. Instead of calculating the average gains of the highest and lowest fourths of a group we may just as well take the ratio of their averages at the beginning of practice and compare it with the ratio of the averages of these same groups of individuals at the end of practice. This measure has the advantage of being more easily calculated.

The reader no doubt has observed that in the above discussion we assumed that the coefficient of variability and that the correlation between initial performance and relative gain are reliable indications of changes in variability at different stages of practice. It is now our task to justify these assumptions and to find out if possible their comparative merit. To do this let us first find an ideal measure of changes in variability.

It seems to us that such a measure would be calculated as follows: First, we calculate the ratios of the best to worst, second best to second worst, third best to third worst, and so on until the ratios of all those above and below the average of the first practice have been found. Second, we calculate the ratios of the members of each pair at the end of practice. Third, we calculate the average of all the ratios at the beginning of practice. Fourth, we may do the same for the final practice. If the average for the fourth step is lower than the average for the third step, the variability of the group has been reduced, while, if this ratio is larger, it has been increased. The objection of this calculation is that it is too long and tedious. We arrive at practically the same result if we calculate the correlation between initial performance and relative gain. For example, if we take the ratio of best to worst at the beginning of practice and find that the ratio for the same individuals at the end of practice has decreased, it means that the worst individual made a larger relative gain than the best. On the other hand, if the ratio has increased, it would mean that the best individual made a larger relative gain than the worst. The same reasoning applies to the ratio of second best to second worst, third best to third worst, etc. In other words, a decrease in the ratios would mean a negative correlation between initial performance and relative gain while an increase in the ratios would mean a positive correlation between initial performance and relative gain. Consequently correlation between initial performance and relative gain is a reliable measure of changes in variability in the course of practice. The coefficient of variability (S. D.÷Ave.) is also a valid measure because it is an approximation to our ideal measure. Instead of being an average of the ratios of pairs of individuals above and below

the mean, it is the ratio of their deviations to the mean, literally $\sqrt{\frac{\Sigma f d^2}{N}}$ . This ratio, however, has a disadvantage in that the numerator represents only that distance from the mean within which the middle two-thirds of the cases fall. It therefore does not adequately represent the whole group, particularly the lowest and the highest sixths. If this reasoning is correct, it follows that Kelley's measure of variability obtained by dividing the difference between the ninetieth and tenth percentiles by the median, $\frac{90\%-10\%}{\text{median}}$, is a better measure than the S. D.÷Average, for the reason that the numerator includes the middle 80 per cent of the cases. However, the disagreements between the two measures would be so rare that it would not pay to calculate both. Because of the limitations mentioned in the S. D.÷Average, this measure will sometimes disagree with the correlation between initial performance and relative gain. Out of 24 experiments, Kincaid found sixteen in which the coefficient of variability showed a decrease but in 22 out of the 24 she found a negative correlation between initial performance and relative gain. There were therefore six cases in which the two measures disagreed, although both of them justify the same general conclusion when we consider the twenty-four experiments as a group. For the reasons stated above the correlation between initial performance and relative gain seems the more valid of the two.

At this point it is necessary to attempt to define the term variability. The term may mean: (1) changes in the output of an individual from one time to another, (2) changes in the average output of a group from one time to another, (3) the amount of scatter or dispersion in the outputs of the individuals of a group, or (4) the ratio of the dispersion of a group to its average performance. The term "individual differences" is also often used to refer to the last two concepts. In this paper we are not interested in the first two usages for which the term,

variation, is a much better word. But the terms, *variability* and *individual differences* as used by the writers whose experiments we have just reviewed is used to cover both definitions (3) and (4); and this confusion is an important cause of their conflicting conclusions. Statisticians also use the word in both of these senses. As measures of variability in the sense of dispersion or scatter, they use the range, the semi-interquartile range or Q, the average deviation or A. D., and the standard deviation or S. D. As a measure of variability in the sense of the ratio of the dispersion of a group to its average performance they use the Q, the A. D. or the S. D. divided by a measure of central tendency, usually the S. D.÷Average, which is called the coefficient of variability. We prefer not to use variability in the sense of the scatter or the amount of the deviations of the individuals from their average. For this we consider the word, dispersion, the best term. Our reason for this is that as achievement is now measured the units of measurement at one stage of practise are not directly comparable with units of measurement at another stage of practice. For example, in typewriting it is possible for some to write 140 words per minute. Practically all persons can make the gain from 0 to 40 words per minute but very few can make the gain from 100 to 140. The last 40 words are therefore thought to be a different amount of gain from the first 40 words. We should be willing to use the term variability in the sense of dispersion if achievement in all practice experiments were measured on a scale starting from an absolute zero so that the units in one part were directly comparable with the units in any other part, but to do so with present methods of measurement would be very misleading for it is entirely possible for a practice group to increase its dispersion and yet reduce its variability. As pointed out above, the students in Kincaid's experiment in Braille writing increased their dispersion as measured by the S. D. from 29 to 92 but reduced their variability from .29 to .17 as measured by the coefficient of variability. In this case, variability means the ratio of dispersion to central tendency or average performance. Such ratios are comparable with each other but the S. D.'s are not. This comparability is

based on the S. D., which is one fraction of the average performance at the beginning of practice and another fraction of the average performance at the end of practice. If during the course of practice this fraction increases, we say that the variability of the group has increased, and if it decreases we say that the variability has decreased. But we also speak of the variability of certain individuals of a group, such as the variability of the highest and lowest individuals or the variability of sections of a group, such as, the variability of the highest and lowest fourths. Because of the incomparability of the units in present methods of measuring achievement, we cannot measure the influence of practice on the variability of highest and lowest individuals by comparing the difference in the raw scores at the beginning of practice with their difference at the end of practice. To make their scores at the end of practice comparable with those at the beginning of practice we calculate their deviations from each average in terms of S. D. units or so-called z scores and then compare the difference in the z scores of the highest and lowest individuals at the beginning of practice with the difference in the z scores of these same individuals at the end of practice. Reduced to mathematical terms this process is as follows:

Let $X_1$ = score of highest individual at beginning of practice
$X_2$ = score of same individual at end of practice
$Y_1$ = score of lowest individual at beginning of practice
$Y_2$ = score of same individual at end of practice
$M_1$ = average of group at beginning of practice
$M_2$ = average of group at end of practice
S.D.$_1$ = standard deviation of $M_1$
S.D.$_2$ = standard deviation of $M_2$

Then in order to compare the variability of these individuals we may calculate $\left(\frac{X_1-M_1}{S.\ D._1}-\frac{Y_1-M_1}{S.\ D._1}\right) - \left(\frac{X_2-M_2}{S.\ D._2}-\frac{Y_2-M_2}{S.\ D._2}\right)$ or instead of taking the difference between these expressions we may calculate their ratios. This is a complicated way of finding the relative differences of the two individuals at the beginning and at end of practice, which may be proved much more simply by comparing $X_1/Y_1$ with $X_2/Y_2$. If we generalize from these cases we may define variability when applied to arbitrary scales

as a certain kind of ratio. When we ask if practice increases or decreases the variability of a group we mean, does it increase or decrease the ratio of dispersion to central tendency? When we ask if practice increases or decreases the variability between the highest and lowest individuals of a group we mean, does it increase or decrease the ratio of their output? And when we ask if practice reduces the variability of certain sections of a group, for example, the lowest and highest fourths, we mean, does it increase or decrease the ratio of their average output? We use ratios rather than absolute scores because ratios are comparable and absolute scores are not, since, in this discussion, we make a sharp distinction between variability and dispersion. However, we believe that these concepts are in need of further analysis before their definitions become final.

If our reasoning on the validity of measures of changes in variability during the course of practice is correct, it follows that in order to get completely reliable results we should use the three following measures: The ratio of the best to the worst at the beginning of practice compared with the ratio of the same individuals at the end of practice, the coefficient of variability (S. D.÷Average), and the correlation between initial performance and relative gain. If we use only one it should be the last. But the first two taken together make an excellent pair, for the first one tells us the changes in variability in the extremes while the second tells us what happens to the middle two-thirds of the group. The first one does not tell us what happens to the group as a whole.

The following measures have a fair reliability: Negative correlation between initial performance and gross gain, the percentage of gain in the upper fourth as compared with that in the lower fourth when the individuals in the two sections are kept the same, the ratio of the average of the highest to the lowest fourths, or of the seventh to the first octiles at the beginning of practice compared with the ratio of the same individuals at the end of practice.

The following have little or no validity as measures of changes in variability during the course of practice and cannot be made

the basis of judgments of such changes: positive correlation between initial and final performance; positive correlation between initial performance and gross gain; ratio of best to worst, next best to next worst, etc., at the beginning of practice compared with the ratios of best to worst, next best to next worst, etc., at the end of practice; gross gains in amount of work per unit of time or amount of time per unit work of the initially high individuals as compared with the initially low individuals; standard deviations and other like measures when taken alone and apart from their averages or other measures of central tendencies.

## Section 4

### Interpretation of Previous Experiments in the Light of Criticism

If now we review the experimental literature from the standpoint of the validity of the measures that were used we find that practically all of them published before 1923 were based on invalid measures. This includes the following experiments: Thorndike, 1916, multiplication, cancelling 2s, addition, mental multiplication, typewriting; Wells, 1912, addition; Chapman, 1914, addition, mental multiplication, cancellation, and color naming; Starch, 1919, mental multiplication; Starch, 1919, substitution; Henmon, 1920, arithmetic; Thurstone, 1918, telegraphy; Perrin, 1918, analogies, mirror reading; Race, 1922, multiplication, addition, cancellation, and other tests. The result is that the conclusion that the effect of equalizing opportunities is to increase individual differences which Thorndike, Starch, Henmon, Chapman, Thurstone, and others thought to be so securely established is now seen to rest on very insecure foundations, and in all probability, is false. With the downfall of this proposition there also goes its corollary which was brought in as an explanation of the facts, namely, that the important and fundamental cause of individual differences is heredity in comparison with which training and environment are minor factors.

It may be said, however, that this proposition is really not a corollary of the first, even if that were true, but is at best an interesting inference.

Even if the conclusions of the above experiments are false, the experimental findings are just as valid as ever. Fortunately most of the experimenters gave their original scores in detail so that it is possible to go over them and calculate other measures of changes in variability which are thought to be more valid. We have gone over all of these on which we could lay our hands and calculated the following measures so far as the data permitted: Ratio of highest to lowest at the beginning of practice and ratio of same individuals at the end of practice; standard deviation divided by average for initial and final performance and correlation between initial performance and relative improvement. The results are given in Table II which may be taken as a summary of the experimental literature hitherto published. The ratio of the highest to the lowest was usually taken from the average of the three highest and the three lowest. But in case there were ties for any of these positions, these were included. In case the measures were in terms of time per unit of work, the highest means the slowest and the lowest means the fastest. Initial and final mean in nearly all cases the first and last trials of the experiment except when the experimenter gave the average or median of the first and last two, three or four trials as the initial and final performances.

In spite of the fact that the experimenters enumerated in the summary table give many contradictory interpretations of their results, we find that this confusion disappears when the data are treated according to the measures given in Table II, which is easily interpreted. Out of 59 experiments from which we were able to compute the ratios of the highest to the lowest we find that in 95 per cent of the cases they decrease as a result of practice. This shows the effects of practice on the differences between the extremes of a distribution. Out of 70 experiments for which we were able to secure the coefficients of variability we find that in 77 per cent of the cases they decrease as a result of practice.

TABLE II

*Summary of previous investigations relating to effect of practice on changes in variability in achievement. Blank under change indicates decrease in variability, <= increase, and 0 = no change.*

| Date | Experimenter | Test | No. and Kind of Subject | Length of Practice | No. of Periods | Measure | Average 3 Highest / Average 3 Lowest Initial | Final | Change | S. D. Average Initial | Final | Change | r Init. Trial and Per cent Gain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1908 | Thorndike | Mental Mult. | 28 Col. Stds. | .. | 16 | Work | 3.91 | 2.69 | .. | .37 | .42 | < | —.26 |
| 1910 | Thorndike | Addition | 19 Col. Stds. | .. | 7 | Work | 2.17 | 1.41 | .. | .26 | .29 | < | —.22 |
| 1911 | Whitley | Cancellation | 9 Col. Stds. | .. | 20 | Time | 1.93 | 1.54 | .. | .27 | .30 | < | —.33 |
| | | Maze | 9 Col. Stds. | .. | 20 | Time | 1.25 | 1.10 | .. | .30 | .25 | .. | —.42 |
| | | Sorting Cards | 8 Col. Stds. | .. | 20 | Time | 1.62 | 1.43 | .. | .32 | .06 | .. | —.92 |
| | | Weight Discr. | 9 Col. Stds. | .. | 16 | Error | 1.05 | .86 | .. | .21 | .12 | .. | —.65[2] |
| | | Multip. | 9 Col. Stds. | .. | 20 | Time | 2.04 | 1.27 | .. | .55 | .37 | .. | —.33 |
| 1912 | Wells | Addition | 5 Adults | 150′ | 30 | Work | 1.38 | 1.29 | .. | .12 | .13 | < | —.52[1] |
| 1913 | Donovan & Thorndike | Addition | 29 Gr. 4 | 60′ | 30 | Work | 3.54 | 2.42 | .. | .37 | .25 | .. | —.35 |
| 1914 | Hahn and Thorndike | Addition | 23 Gr. 4 | 90′ | 4–6 | Work | 8.33 | 2.69 | .. | .59 | .53 | .. | —.61 |
| | | | 38 Gr. 5 | 90′ | 5–1 | Work | 6.26 | 2.89 | .. | .47 | .46 | .. | —.35 |
| | | | 27 Gr. 6 | 90′ | 5–8 | Work | 9.00 | 8.25 | .. | .46 | .44 | .. | —.41 |
| | | | 26 Gr. 7 | 90′ | 7–11 | Work | 3.21 | 1.91 | .. | .29 | .28 | .. | —.42 |
| 1914 | Brown, W. | Card Sorting | 26 Col. Stds. | .. | 8 | Work | 1.51 | 1.28 | .. | .13 | .10 | .. | ... |
| 1914 | Chapman | Addition | 22 Col. Stds. | 100′ | 10 | Work | 2.23 | 2.20 | .. | .29 | .26 | .. | —.43 |
| | | Cancelling 2 S | 22 Col. Stds. | 10′ | 10 | Work | 1.61 | 1.35 | .. | .13 | .14 | .. | —.24 |
| | | Cancelling 3 S | 22 Col. Stds. | 10′ | 10 | Work | 1.49 | 1.57 | < | .12 | .14 | < | +.15 |
| | | Color Naming | 22 Col. Stds. | 15′ | 10 | Work | 1.52 | 1.22 | .. | .13 | .11 | .. | —.36 |
| | | Opposites | 22 Col. Stds. | 15′ | 10 | Work | 2.53 | 1.16 | .. | .21 | .15 | .. | —.55 |
| 1916 | Thorndike | Addition | 11 Col. Stds. | .. | 10 | Work | 2.10 | 1.64 | .. | .26 | .21 | .. | —.56 |
| | | Cancelling 2 S | 11 Col. Stds. | .. | 10 | Work | 1.30 | 1.05 | .. | .13 | .10 | .. | —.89 |
| | | Cancelling 3 S | 11 Col. Stds. | .. | 10 | Work | 1.47 | 1.39 | .. | .12 | .98 | .. | —.58 |
| | | Mental Mult. | 11 Col. Stds. | .. | 10 | Work | 1.52 | 1.34 | .. | .28 | .40 | < | —.18 |
| | | Typing | 11 Col. Stds. | .. | 10 | Work | 2.22 | 1.62 | .. | .41 | .18 | .. | —.79 |
| | Murphy | Javelin Throwing | 30 Col. Stds. | .. | 100–150 | Errors | 2.85 | 1.75 | .. | .34 | .24 | .. | —.59 |

| | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1916 | McCall | Addition | 72 Chil. | 180′ | 18 | Work | 4.92 | 1.37 | .. | .45 | .51 | < | +.27 |
| | | Cancelling S's | 71 Chil. | 9′ | 9 | Work | 4.43 | 1.85 | .. | .51 | .21 | .. | —.55 |
| | | Cancelling 3 S | 71 Chil. | 16′ | 16 | Work | 3.00 | 1.82 | .. | .30 | .18 | .. | —.81 |
| | | Copying | 75 Chil. | 260′ | 20 | Work | 8.22 | 1.68 | .. | .36 | .23 | .. | —.63 |
| 1917 | Peterson | Ball Tossing | 28 Col. Stds. | .. | 30 | Time | 2.88 | 2.51 | .. | .28 | .40 | .. | —.27[5] |
| 1917 | Woodrow | Form Sorting F.M. | 20 Chil. | 104′ | 13 | Work | 1.24 | 1.52 | < | .14 | .18 | < | —.40 |
| | | Sorting Norman | 16 Chil. | 104 | 13 | Work | 2.12 | 1.31 | .. | .13 | .13 | 0 | —.36 |
| 1918 | Thurstone | Telegraphy | 165 Adults | 72 Hrs. | .. | .... | .... | .... | .. | .72 | .52 | .. | ... |
| 1918 | Myers | Card Sorting | 7 Col. Stds. | .. | 23 | Time | .... | .... | .. | .08 | .13 | < | ... |
| | | Cube Fitting | 7 Col. Stds. | .. | 23 | Time | .... | .... | .. | .12 | .12 | 0 | ... |
| 1919 | Perrin | Analogies | 21 Col. Stds. | .. | 10 | Time | 4.88 | 1.42 | .. | .53 | .41 | .. | —.52 |
| | | Mir. Reading | 21 Col. Stds. | .. | 7 | Time | 2.46 | 1.71 | .. | .60 | .50 | .. | —.44 |
| 1919 | Starch | Mental Mult. | 8 Col. Stds. | .. | 14 | Work | 1.56 | 1.64 | < | ... | ... | .. | ... |
| | | Substitution | 20 Col. Stds. | 120 | 10 | Work | 1.39 | 1.26 | .. | ... | ... | .. | ... |
| 1919 | Chapman | Typing | 20 H.S. Stds. | 180 | 180 | Work | 8.13 | 4.33 | .. | .38 | .11 | .. | —.41 |
| 1922 | Race | Addition Ex. A. & B. | 59 Col. Stds. | 160′ | 8 | Work | 4.68 | 2.56 | .. | .39 | .31 | .. | —.43 |
| | | Addition | 71 Aver. Chil. | 80′ | 10 | Work | 15.15 | 4.71 | .. | .56 | .51 | .. | —.43 |
| | | Addition | 33 Super. Chil. | 80′ | 10 | Work | 6.23 | 3.25 | .. | .49 | .49 | 0 | —.31 |
| | | Cancellation | 59 Col. Stds. | 160′ | 8 | Work | 1.80 | 1.52 | .. | .15 | .13 | .. | —.53 |
| | | Ex. A. & B. History | 33 Super. Chil. | 800′ | 20 | Work | 2.93 | 2.00 | .. | .27 | .20 | .. | —.03 |
| | | History | 25 Gr. 7 | 800′ | 20 | Work | 3.22 | 1.50 | .. | .34 | .24 | .. | —.55 |
| | | History | 39 Gr. 5 | 800′ | 20 | Work | 7.61 | 3.95 | .. | .42 | .41 | .. | —.35 |
| | | Intel. Tests | 38—19 Sup. 19 Ave. | 450′ | 15 | Work | 2.95 | 1.81 | .. | .28 | .18 | .. | —.44 |
| | | Language | 31 Super. Chil. | 100′ | 10 | Work | 1.47 | 1.23 | .. | .11 | .09 | .. | —.49 |
| | | Multip. Exp. A.&B. | 59 Col. Stds. | 160′ | 8 | Work | 4.46 | 1.32 | .. | .32 | .22 | .. | —.67 |
| | | Multip. | 34 Super. Chil. | 160′ | 20 | Work | 2.87 | 1.97 | .. | .18 | .03 | .. | +.03 |
| | | Multip. | 52 Aver. Chil. | 160′ | 20 | Work | 2.51 | 1.69 | .. | .25 | .22 | .. | —.54 |
| 1922 | Gates, G. S. | Addition | 23 Col. Stds. | .. | 29 | Work | 1.63 | 1.36 | .. | .16 | .12 | .. | +.16[2] |
| | | Color Naming | 23 Col. Stds. | .. | 22 | Work | 1.65 | 1.43 | .. | .21 | .17 | .. | —.24[2] |
| | | Mental Mult. | 23 Col. Stds. | .. | 22 | Work | 4.08 | 3.61 | .. | .45 | .43 | .. | —.37[2] |
| | | Tapping | 23 Col. Stds. | .. | 22 | Work | 1.32 | 1.02 | .. | .10 | .12 | < | —.15[2] |

TABLE II—Continued

| Date | Experimenter | Test | No. and Kind of Subject | Length of Practice | No. of Periods | Measure | Average 3 Highest / Average 3 Lowest Initial | Final | Change | S. D. Initial | Final | Change | r Init. Trial and Per cent Gain |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1924 | Reed | Addition | 140 Col. Stds. | 60′ | 1–6 | Work | 2.82 | 1.93 | .. | .80 | .21 | .. | —.75 |
| | | Alphabet Back. | 129 Col. Stds. | 12 Rep. | 1 | Time | 11.30 | 2.42 | .. | .59 | .42 | .. | —.56 |
| | | Substitution | 108 Col. Stds. | 15 Stars | 1 | Time | 12.65 | 2.20 | .. | .53 | .56 | < | —.36 |
| | | Mirror Drawing | 58 Col. Stds. | 400 Subs. | 1 | Time | 4.20 | 0.94 | .. | .32 | .24 | .. | —.65 |
| | | Substitution | 59 Col. Stds. | 400 Subs. | 1 | Work | 3.22 | 2.49 | .. | .43 | .42 | .. | —.23 |
| 1924 | Barlow | Card Sorting | 20 Chil. | 720 Sorts | 6 | Time | .... | .... | .. | .15 | .12 | .. | ... |
| | | Mental Mult. | 20 Chil. | 240–400 | 12–20 | Work | .... | .... | .. | .28 | .34 | < | ... |
| | | Substitution | 20 Chil. | 2400 | 12 | Time | .... | .... | .. | .15 | .18 | < | ... |
| 1925 | Kincaid | Braille Writing | 31 Col. Stds. | .. | 47 | ..... | .... | .... | .. | .29 | .17 | .. | —.89 |
| | | Dart Throwing | 28 Col. Stds. | .. | 40 | ..... | .... | .... | .. | .38 | .31 | .. | —.60 |
| 1928 | Peterson & Barlow | Mental Multip. | 96 Chil. | 882 Prob. | 7 | Time | .... | .... | .. | .37[3] | .32[3] | .. | ... |
| | | Psychology | 56 Col. Stds. | 10 Wks. | .. | Work | .... | .... | .. | .29[4] | .17[4] | .. | ... |
| | | Psychology | 56 Col. Stds. | 9 Wks. | .. | Work | .... | .... | .. | .35[4] | .11[4] | .. | ... |
| | | Psychology | 56 Col. Stds. | 8 Wks. | .. | Work | .... | .... | .. | .23[4] | .11[4] | .. | ... |
| | | Psychology | 56 Col. Stds. | 10 Wks. | .. | Work | .... | .... | .. | .33[4] | .21[4] | .. | ... |
| 1929 | Sandiford | Typing | 10 H.S. Stds. | 360 Hrs. | .. | Work | 2.07 | 1.10 | .. | .09 | .02 | .. | —.98[2] |
| Number of experiments from which measurements were calculated | | | | | | | | 59 | | | 70 | | 58 |
| Per cent showing decrease in variability | | | | | | | | 95 | | | 77 | | 93 |
| Per cent showing increase in variability | | | | | | | | 5 | | | 19 | | 7 |
| Per cent showing no change in variability | | | | | | | | 0 | | | 4 | | 0 |

The coefficients of variability and the correlations for Thorndike, 1908; Thorndike, 1910; Whitley, 1911, multiplication, cancellation, and maze; Thorndike, 1916; Chapman, 1914; McCall, 1916; and Kincaid, 1925, were taken from Kincaid's table. The coefficients of variability for Brown, 1914, and Barlow, 1924, were taken from Barlow's table. The coefficients of variability for Thurstone, 1918, and Peterson, 1928, were taken from Peterson's table. 1 = R. 2 = Rho. $3 = \frac{90P—10P.}{2Md.}$ 4 = Q/Md. 5. Peterson's results were reduced to time by counting each error as 5 seconds and each catch as 1 second.

This shows the effect of practice on the middle two-thirds of a distribution. Out of 58 experiments for which we were able to get the correlation between initial performance and percentage of improvement we find that 93 per cent of them are negative. This shows the effect of practice on the group as a whole. In other words, practice nearly always decreases the variability in the achievements of a group of individuals. This seems to be true for all kinds of tests, mental or motor, simple or complex. There are, however, differences in the amount of decrease. An inspection of the table indicates that this is influenced by three factors, namely, the amount of difference between individuals, the length of the practice, and the initial efficiency of the individuals. The greater the difference between two individuals the greater the decrease; the longer the practice the greater the decrease; and the higher the initial efficiency, the less the decrease. All of this suggests that the limits of improvement set by uncontrollable factors are much closer together than those due to training, and as individuals approach these limits because of equal opportunity and training, their relative differences in achievement are decreased.

# CHAPTER II

# NEW EXPERIMENTS

## Section 1

### An Experiment with the Yo-Yo Top

We now wish to report two new investigations on the influence of practice on changes in variability in achievement—one a motor test and the other a series of tests of achievement in school subjects. The motor test was done by a group of thirty-seven college sophomores and consisted of one hundred eighty minutes practice in throwing the yo-yo top. The task was to throw the top down and allow it to roll back on the string into the hand. If it did, such a catch was considered one success. The number of catches in ten minutes was the score. The students were divided into three groups which distributed the time in three ways. One group practiced ten minutes a day for eighteen days; another, twenty minutes a day for nine days; and a third group, sixty minutes once a week for three weeks. In calculating the influence of practice on changes in variability no account was taken of the effect of the distribution of practice. Table III gives the results for the first and eighteenth periods of practice.

In order to show the effect of different measures we calculated the following:

1. The averages.
2. The standard deviations.
3. The coefficients of variability.
4. The correlation between initial and final performance.
5. The correlation between initial performance and gross gain.
6. The correlation between initial performance and relative gain.

TABLE III

*Results from tossing yo-yo top. Columns one and two show number of catches in ten minutes*

| First performance | Eighteenth performance | Gross gain | Per cent gain |
|---|---|---|---|
| 2 | 262 | 260 | 13000 |
| 2 | 530 | 528 | 26400 |
| 3 | 412 | 409 | 13633 |
| 4 | 145 | 141 | 3525 |
| 4 | 506 | 502 | 12550 |
| 5 | 401 | 396 | 7920 |
| 7 | 353 | 346 | 4943 |
| 16 | 523 | 507 | 3168 |
| 19 | 420 | 401 | 2110 |
| 20 | 312 | 492 | 2460 |
| 29 | 415 | 386 | 1331 |
| 29 | 419 | 390 | 1348 |
| 31 | 398 | 367 | 1184 |
| 40 | 402 | 362 | 905 |
| 45 | 520 | 470 | 1044 |
| 48 | 448 | 400 | 833 |
| 49 | 373 | 324 | 661 |
| 50 | 331 | 281 | 562 |
| 55 | 500 | 445 | 809 |
| 66 | 249 | 183 | 277 |
| 68 | 400 | 332 | 488 |
| 70 | 575 | 505 | 721 |
| 74 | 350 | 276 | 373 |
| 93 | 481 | 388 | 417 |
| 96 | 600 | 504 | 525 |
| 100 | 553 | 453 | 453 |
| 101 | 535 | 434 | 429 |
| 120 | 275 | 155 | 129 |
| 123 | 519 | 396 | 322 |
| 134 | 359 | 225 | 167 |
| 138 | 538 | 400 | 2107 |
| 147 | 608 | 461 | 313 |
| 172 | 403 | 231 | 134 |
| 198 | 642 | 444 | 224 |
| 235 | 382 | 147 | 62 |
| 260 | 420 | 160 | 61 |
| 400 | 660 | 260 | 65 |
| Averages 82.51 | 445.63 | 852.18 | 28.06 |
| S.D. 84.13 | 112.72 | 112.19 | 53.25 |
| V or S.D./Ave. 1.02 | .25 | | |

$V_1-V_2=.77$. P.E. diff. $=.17$.

TABLE IV

| | Correlations |
|---|---|
| Initial and final performance | .37 |
| Initial performance and gross gain | .—62 |
| Initial performance and per cent gain | .—43 |

TABLE V

| Ratios | Initial | Final |
|---|---|---|
| 3 Highest / 3 Lowest | 99.44 | 1.31 |
| 7th Octile / 1st Octile | 46.25 | 1.39 |
| 3rd Quartile / 1st Quartile | 6.28 | .85 |
| 3 Lowest / 4 Middle | 27.78 | .95 |

TABLE VI

*Averages of certain groups*

| | Initial score | Final score | Gross gain | Relative gain |
|---|---|---|---|---|
| 3 Highest | 298.3 | 487.33 | 189. | 63 |
| 7th Octile | 185. | 522.5 | 337.5 | 182 |
| 3rd Quartile | 122.5 | 397. | 274.5 | 224 |
| 4 Middle | 83.25 | 501.5 | 418.2 | 502 |
| 1st Quartile | 19.5 | 466. | 466. | 2290 |
| 1st Octile | 4. | 385.5 | 371.5 | 9288 |
| 3 Lowest | 3. | 371. | 368. | 12267 |

TABLE VII

*Averages of certain groups, gross and relative gains on basis of time per 10,000 catches*

| Min. per 10,000 | Initial | Final | Gross gain | Relative gain |
|---|---|---|---|---|
| 3 Highest | 335 | 205 | 130 | 38.8 |
| 7th Octile | 540 | 191 | 349 | 64.6 |
| 3rd Quartile | 816 | 252 | 564 | 69.1 |
| 4 Middle | 1201 | 199 | 1002 | 83.4 |
| 1st Quartile | 5025 | 214 | 4811 | 95.7 |
| 1st Octile | 25000 | 266 | 24731 | 98.9 |
| 3 Lowest | 33333 | 269 | 33064 | 99.1 |

7. The ratio of the average of the three highest to that of the three lowest in the initial practice and the ratio of the same individuals in the final practice.

8. The ratio of the average of two on the seventh octile to the two on the first octile in the initial practice and the ratio of the same individuals in the final practice.

9. The ratio of the average of two scores on the third quartile to the two on the first quartile in the first practice and the ratio of the same individuals in the last practice.

10. The ratio of average of four middle scores of the initial practice to that of the three lowest, and the ratio of the same individuals in the final practice.

11. The gross and relative gains of the following groups as defined above.

a. The three highest
b. the seventh octile
c. the third quartile
d. the four middle individuals
e. the first quartile
f. the three lowest.

12. The gross and relative gains of the group named in No. 11 but calculated on the basis of amount of time per unit of work. This was done by dividing the scores on the basis of amount of work per unit of time into ten and then multiplying the resulting quotient by ten thousand in order to avoid decimals.

The results of the above calculations are given in Tables III to VII. In interpreting them we see first that there was a very great improvement which is indicated by the average increasing from 82.51 to 445.63. If we based our judgment of the effect of changes in variability on the S. D.'s alone, we would say there was a great increase, but as S. D.'s have no meaning apart from their averages, we notice that when taken in relation to them as in the coefficient of variability, there is a decrease from 1.02 to .25. The positive correlation of .37 indicates that there is a general tendency for individuals of the group to maintain their ranks but the negative correlation of —.62 between initial performance and gross gain and the one of —.43 between initial performance and relative gain shows that the lower the initial score the greater the gain will be both absolutely and relatively. Turning to the Table V on ratios we see that they are very large in the first practice, varying from 99.44 between the three highest and the three lowest to 6.28 between the third and first quartiles, but in the eighteenth practice, they have all been reduced to less than 1.40. It is particularly striking to notice that while in the first practice the three highest were nearly one hundred times

as efficient as the three lowest, according to the scale used, they were less than twice as efficient as the three lowest after 180 minutes of practice. The reason for this decrease is that the three highest were already fairly efficient with the yo-yo when the systematic practice began but the three lowest never had one in their hands. However after they had a fair opportunity to learn, they were not far behind the best ones. The ratios also inform us that the farther apart individuals are in a skill before having an equal opportunity to acquire it, the greater their convergence will be after taking advantage of such an opportunity.

Turning now to Table VI which shows the gross and relative gains of groups according to their rank from highest to lowest, we notice a striking inverse relationship between initial score and percentage of gain which, like the results from the experiments hitherto considered, shows that practice reduces variability, but when we study the gross gains, this conclusion is not so evident. In fact from the three lowest to the four middle individuals we see that with a single exception the higher the initial score the greater the gain, but the scores from the middle to the three highest indicate just the opposite. If by chance our group had been such as to exclude its upper half and if we had restricted ourselves to gross gains in amount of work per unit of time as a measure of the influence of practice on variability we could have presented our results as in Table VIII.

TABLE VIII

*Gains in amount of work per unit of time in relation to initial score*

| | Average | Gain |
|---|---|---|
| Initially 4 highest | 83.25 | 418.2 |
| Initially 3 lowest | 3. | 368. |

The facts so arranged warrant the same conclusion that Thorndike drew from one of his experiments in which gross gains were used, namely: "In general, the gross improvement in product per unit of time is greatest for those of initial highest ability. The effect of equalizing opportunity is to leave the gross variability between individuals unimpaired or even to increase it."

If now we convert these scores into time per unit of work, we have the facts as in Table IX.

TABLE IX

*Gains in amount of time per unit of work in relation to initial score*

| | Average initial score | Gain |
|---|---|---|
| Initially 4 highest | 1201 | 1002 |
| Initially 3 lowest | 33,333 | 33,064 |

From the facts as now arranged we could conclude:

"In general the gross improvement in time per unit of product is greatest for those of initial lowest ability. The effect of equalizing opportunity is to leave the gross variability of individuals greatly reduced."

These contradictory conclusions from the same facts measured in two equally good ways are presented to again show in a concrete way that judgments about changes in variability as a result of practice cannot be based on absolute scores either in terms of amount of work per unit of time or amount of time per unit of work. This is apparent not only from the opposite conclusions to which they lead but also from the relative gains shown in Tables VI and VII. These, whether based on amount of work per unit of time or amount of time per unit of work, show consistently that the effect of practice is to reduce variability, a judgment which is also justified by the coefficient of variability, the negative correlations between gain and initial score, and the various ratios employed.

## Section 2

### An Experiment on Influence of Public School Training on Variability in Achievement

The investigation with a series of tests of achievement in school subjects was an attempt to discover what changes in variability in achievement in school subjects result from eight months training as it is usually given in the public schools. This work was done under our supervision by Mr. C. J. Edwards, Super-

intendent of Schools, Elizabeth, Colorado, as a partial fulfillment of the requirement for the master's degree.

The materials for this investigation consisted of the results obtained from three surveys made of a small public school system early in October, 1928, early in February, 1929, and late in May, 1929. The purpose of giving these tests was supervision alone, but after the papers were scored twice, and the scores carefully transferred to large sheets of paper for convenience of filing and interpretation, we saw that here was good material for the research problem under discussion and we therefore used it for this purpose also. All of the tests were given, scored, and recorded under our supervision. They consisted of the Stanford Achievement Tests, The Compass Survey Arithmetic Tests, and some others, but only the Stanford and Compass tests were used for this investigation. The Stanford Primary Examination was used in Grades II and III and the Advanced Examination in the remaining grades. The Compass Primary Examination was used in Grades II, III, and IV, and the Advanced Examination in the remaining grades.

The Stanford Achievement Test and Primary Examination consists of the following tests.

1. Reading: Paragraph Meaning.
2. Reading: Sentence Meaning.
3. Reading: Word Meaning.
4. Arithmetic Computation.
5. Arithmetic Reasoning.
6. Dictation.

The Stanford Achievement Test Advanced Examination includes the following tests:

1. Reading: Paragraph Meaning.
2. Reading: Sentence Meaning.
3. Reading: Word Meaning.
4. Arithmetic Computation.
5. Arithmetic Reasoning.

6. Nature Study and Science.
7. History and Literature.
8. Language Usage.
9. Dictation Exercises.

The Compass Survey Arithmetic Tests Elementary Examination includes addition, subtraction, multiplication and division. The Advanced Examination includes all of these and also percentage and general problems.

For the purposes of this research only the records of children were used who took the three survey tests in October, February and May. This included twenty in Grade 2, twenty in Grade 3, twenty-two in Grade 5, seventeen in Grade 4, nineteen in Grade 6, nineteen in Grade 7, and twenty in Grade 8, or a total of one hundred thirty-seven pupils.

Mr. Edwards used the following measures of variability:

1. Ratio of average of the three highest scores to the average of three lowest in the October tests. In the February and May tests the ratios of the averages of the same pupils were calculated. The results of this measure are given in Tables X and XI.

2. The coefficient of variability or standard deviation divided by the mean. The results of this measure are given in Tables XII and XIII.

3. The ratio of the seventy-fifth percentile to the twenty-fifth percentile in the October tests. For the February and May tests the ratios of the scores of the same pupils were calculated. The results of this measure are given in Tables XIV and XV.

4. The ratio of the sixtieth percentile to the fortieth percentile in the October tests. In the February and May tests, the ratios of the scores of the same pupils were used. The results of this measure are given in Tables XVI and XVII.

5. The correlation between the October scores in each test and per cent of gain from October to May. The results of this measure are given in Table XVIII.

The first four measures were calculated for the October, February, and May scores on each test for the pupils of each grade, but for the fifth measure, the grade divisions were ignored and the scores assembled by tests. For example, in calculating the correlation between the October scores and the per cent of gain from October to May for Test I, Reading: Paragraph Meaning, the scores of all the pupils from Grades II to VIII inclusive were used.

In interpreting the results we shall find it convenient to first look at the averages and then at the cases from which the averages are computed. Let us first look at Table XI which presents the average by grades of the ratios of the highest to the lowest. The October average of 8.19 for Grade IV is the average of ten ratios, one for each of the tests given in this grade. The other averages in the tables were similarly derived. Comparing the October and May averages, we notice that every pair of ratios shows a marked decrease from October to May. The same is true from October to February but from February to May there is one case, Grade VI, that shows a small increase. The interpretation of this measure is clear from the final averages, which show a decrease from 6.96 to 1.75 from October to February and a decrease from 6.96 to 1.68 from October to May. If we look at the individual ratios in Table XI we find that 98 per cent of them show a decrease from October to May. From this it clearly follows that the differences in the achievements of the highest and lowest of a group are much reduced by equal training.

TABLE X

*Grade averages of ratios of average of 3 highest to average of lowest in each test for October, February, and May. From Edwards, 1929*

| | October | February | May |
|---|---|---|---|
| Grade II | 10.94 | 1.93 | 1.60 |
| Grade III | 16.97 | 1.91 | 1.60 |
| Grade IV | 8.19 | 1.76 | 1.74 |
| Grade V | 4.98 | 2.00 | 1.87 |
| Grade VI | 2.73 | 1.41 | 1.82 |
| Grade VII | 2.41 | 1.47 | 1.30 |
| Grade VIII | 2.54 | 1.81 | 1.87 |
| Final Average | 6.96 | 1.75 | 1.68 |

TABLE XI

*Ratios of average of 3 highest to average of 3 lowest for each test and grade. From Edwards, 1929*

| | October | February | May |
|---|---|---|---|
| **GRADE II** | | | |
| Total Reading | 26.25 | 3.25 | 2.03 |
| Total Arithmetic | 1.85 | 1.25 | 1.31 |
| Compass Arithmetic | 4.73 | 1.31 | 1.48 |
| **GRADE III** | | | |
| Test | | | |
| I | 66.66 | 1.83 | 1.82 |
| II | 27. | 2.65 | 1.74 |
| III | 8.10 | 2.00 | 1.18 |
| IV | 4.11 | 1.42 | 1.13 |
| V | 4. | 1.52 | 1.94 |
| VI | 6.33 | 1.96 | 1.67 |
| VII | 2.60 | 2.03 | 1.75 |
| **GRADE IV** | | | |
| Test | | | |
| I | 5.03 | 2.31 | 1.85 |
| II | 4.74 | 1.29 | 1.70 |
| III | 4.72 | 1.23 | 1.37 |
| IV | 2.33 | 1.39 | 1.25 |
| V | 5.85 | 1.54 | 1.88 |
| VI | 21.30 | 4.46 | 2.93 |
| VII | 28.27 | 0.00 | 2.08 |
| VIII | 6.15 | 2.97 | 1.59 |
| IX | 1.87 | 1.38 | 1.69 |
| Com. Arith. | 1.68 | 1.09 | 1.15 |
| **GRADE V** | | | |
| Test | | | |
| I | 3.08 | 1.75 | 1.66 |
| II | 5.44 | 2.70 | 1.88 |
| III | 2.45 | 1.83 | 1.98 |
| IV | 1.75 | 1.15 | 1.17 |
| V | 3.28 | 2.00 | 2.10 |
| VI | 4.37 | 2.51 | 2.44 |
| VII | 9.02 | 2.60 | 1.64 |
| VIII | 13.40 | 1.75 | 2.14 |
| IX | 2.29 | 2.07 | 1.83 |
| Com. Arith. | 4.73 | 1.66 | 1.89 |
| **GRADE VI** | | | |
| Test | | | |
| I | 2.62 | 1.58 | 2.10 |
| II | 2.60 | 1.51 | 2.57 |
| III | 1.83 | 1.55 | 1.36 |
| IV | 1.79 | .97 | 1.07 |
| V | 2.07 | 1.09 | 1.52 |
| VI | 2.47 | 1.48 | 1.98 |
| VII | 7.35 | 1.09 | 2.28 |
| VIII | 1.94 | 1.49 | 2.04 |
| IX | 1.45 | 1.45 | 1.62 |
| Com. Arith. | 3.27 | 1.92 | 1.75 |

TABLE XI—Continued

GRADE VII

| Test | October | February | May |
|---|---|---|---|
| I | 1.65 | 1.28 | 1.17 |
| II | 1.92 | 1.46 | 1.35 |
| III | 1.62 | 1.48 | 1.48 |
| IV | 1.50 | 1.59 | 1.20 |
| V | 1.97 | 1.57 | 1.45 |
| VI | 1.41 | 1.40 | 1.21 |
| VII | 3.21 | 1.61 | 1.28 |
| VIII | 5.30 | 1.23 | 1.13 |
| IX | 1.65 | 1.41 | 1.45 |
| Com. Arith. | 3.87 | 1.75 | 0.00 |

GRADE VIII

| Test | | | |
|---|---|---|---|
| I | 2.09 | 1.55 | 1.54 |
| II | 3.51 | 2.78 | 2.92 |
| III | 1.80 | 1.65 | 1.75 |
| IV | 1.41 | 1.31 | 1.30 |
| V | 2.53 | 1.94 | 1.91 |
| VI | 2.49 | 1.48 | 1.51 |
| VII | 4.52 | 2.43 | 3.01 |
| VIII | 3.26 | 1.68 | 2.13 |
| IX | 1.54 | 1.25 | 1.26 |
| Com. Arith. | 2.33 | 2.04 | 1.45 |

If now we turn to Table XII which presents the grade averages of the coefficients of variability we see what changes in variability occur primarily in the central two-thirds of the groups.

TABLE XII

*Grade averages of coefficients of variability of each test for October, February and May*

| Grade | October | February | May |
|---|---|---|---|
| II | 1.02 | .57 | .37 |
| III | .53 | .29 | .28 |
| IV | .44 | .32 | .28 |
| V | .37 | .32 | .28 |
| VI | .26 | .23 | .26 |
| VII | .24 | .22 | .18 |
| VIII | .24 | .23 | .21 |
| Final Average | .45 | .28 | .26 |

TABLE XIII

*Coefficients of variability (S. D.:Ave.) for each test and grade for October, February and May. From Edwards, 1929*

| Test | October | February | May |
|---|---|---|---|
| **Grade II** | | | |
| I | 1.55 | .95 | .50 |
| II | 1.38 | .69 | .52 |
| III | .90 | .62 | .59 |
| IV | .37 | .40 | .18 |
| V | 1.41 | .68 | .27 |
| VI | 1.06 | .34 | .34 |
| Com. Arith. | .48 | .33 | .19 |
| Average | 1.02 | .57 | .37 |
| **Grade III** | | | |
| I | .72 | .35 | .34 |
| II | .64 | .43 | .34 |
| III | .57 | .46 | .32 |
| IV | .48 | .20 | .21 |
| V | .45 | .37 | .36 |
| VI | .61 | .26 | .20 |
| Com. Arith. | .27 | .27 | .25 |
| Average | .53 | .29 | .28 |
| **Grade IV** | | | |
| I | .39 | .27 | .22 |
| II | .45 | .22 | .33 |
| III | .45 | .24 | .26 |
| IV | .25 | .15 | .15 |
| V | .52 | .32 | .31 |
| VI | .63 | .36 | .38 |
| VII | .76 | .65 | 1.06 |
| VIII | .57 | .66 | .40 |
| IX | .20 | .18 | .18 |
| Com. Arith. | .18 | .15 | .15 |
| Average | .44 | .32 | .34 |
| **Grade V** | | | |
| I | .31 | .22 | .22 |
| II | .37 | .42 | .32 |
| III | .26 | .25 | .26 |
| IV | .17 | .11 | .13 |
| V | .32 | .22 | .24 |
| VI | .39 | .28 | .28 |
| VII | .53 | .58 | .37 |
| VIII | .60 | .60 | .43 |
| IX | .28 | .30 | .28 |
| Com. Arith. | .53 | .29 | .28 |
| Average | .37 | .32 | .28 |

TABLE XIII—Continued

GRADE VI

| Tests | October | February | May |
|---|---|---|---|
| I | .25 | .16 | .22 |
| II | .28 | .21 | .32 |
| III | .17 | .19 | .17 |
| IV | .20 | .14 | .15 |
| V | .22 | .22 | .29 |
| VI | .26 | .23 | .26 |
| VII | .57 | .46 | .36 |
| VIII | .22 | .32 | .36 |
| IX | .13 | .17 | .18 |
| Com. Arith. | .37 | .27 | .31 |
| Average | .26 | .23 | .26 |

GRADE VII

| Test | | | |
|---|---|---|---|
| I | .16 | .14 | .12 |
| II | .22 | .26 | .20 |
| III | .17 | .16 | .13 |
| IV | .12 | .22 | .16 |
| V | .23 | .19 | .20 |
| VI | .12 | .19 | .14 |
| VII | .34 | .39 | .30 |
| VIII | .41 | .25 | .25 |
| IX | .19 | .19 | .15 |
| Com. Arith. | .51 | .21 | .00 |
| Average | .24 | .22 | .18 |

GRADE VIII

| Test | | | |
|---|---|---|---|
| I | .21 | .17 | .16 |
| II | .31 | .29 | .27 |
| III | .19 | .21 | .20 |
| IV | .19 | .12 | .14 |
| V | .26 | .22 | .21 |
| VI | .27 | .23 | .21 |
| VII | .34 | .38 | .36 |
| VIII | .33 | .33 | .33 |
| IX | .14 | .15 | .13 |
| Com. Arith. | .23 | .20 | .16 |
| Average | .24 | .23 | .21 |

Table XII shows that without exception the coefficients show a marked decrease both from October to February and from October to May. There is one case, Grade VI, which shows increase from February to May. The final averages which show a decrease from .45 to .26 from October to May indicate the changes that occur according to this measure. The sixty-four rows of coefficients in Table XIII, from which the averages

in Table XII were calculated, show that 76 per cent of them decrease from October to February and that 85 per cent of them decrease from October to May. From this it follows that although the variability of the central two-thirds of a group does not decrease so universally as does that of the extremes, yet in the large majority of cases the differences between the members of this portion of the group also decrease from the same training.

TABLE XIV

*Grade averages of the ratio of the upper quartile to lower quartile for October, February and May*

| Grade | October | February | May |
|---|---|---|---|
| II | 2.31 | 1.19 | 1.56 |
| III | 2.75 | 1.49 | 1.59 |
| IV | 2.13 | 1.03 | 1.17 |
| V | 1.72 | 1.19 | 1.15 |
| VI | 1.72 | 1.26 | 1.35 |
| VII | 1.40 | 1.30 | 1.33 |
| VIII | 1.46 | 1.30 | 1.00 |
| Final Average | 1.93 | 1.25 | 1.31 |

TABLE XV

*Ratio of upper quartile to lower quartile in each test in each grade for October, February and May*

GRADE II

| | October | February | May |
|---|---|---|---|
| Total Reading | 2.66 | 1.11 | 2.33 |
| Total Arithmetic | 1.50 | .94 | 1.01 |
| Compass Arithmetic | 2.75 | 1.54 | 1.33 |
| Average | 2.31 | 1.19 | 1.56 |

GRADE III

| Test | | | |
|---|---|---|---|
| I | 6.00 | 2.18 | 2.10 |
| II | 2.83 | .58 | 1.23 |
| III | 2.85 | 2.25 | 2.50 |
| IV | 1.66 | 1.36 | 1.25 |
| V | 2.00 | 1.23 | 1.20 |
| VI | 2.37 | 1.48 | 1.11 |
| Comp. Arith. | 1.55 | 1.34 | 1.74 |
| Average | 2.75 | 1.49 | 1.59 |

TABLE XV—Continued

GRADE IV

| Test | October | February | May |
|---|---|---|---|
| I | 1.38 | 1.31 | 1.21 |
| II | 1.94 | 1.31 | 1.00 |
| III | 2.14 | .87 | .82 |
| IV | 1.33 | 1.05 | 1.50 |
| V | 2.00 | 1.30 | 2.16 |
| VI | 2.44 | .93 | 1.61 |
| VII | 4.66 | .88 | 1.04 |
| VIII | 2.71 | .81 | .14 |
| IX | 1.28 | .85 | 1.18 |
| Com. Arith. | 1.43 | 1.03 | 1.05 |
| Average | 2.13 | 1.03 | 1.17 |

GRADE V

| Test | | | |
|---|---|---|---|
| I | 1.47 | 1.27 | 1.15 |
| II | 1.32 | 1.37 | 1.40 |
| III | 1.44 | .96 | 1.24 |
| IV | 1.31 | 1.08 | 1.13 |
| V | 1.75 | 1.18 | 1.25 |
| VI | 1.90 | 1.30 | 1.57 |
| VII | 2.62 | .79 | .60 |
| VIII | 2.33 | 1.46 | .55 |
| IX | 1.37 | 1.42 | 1.36 |
| Com. Arith. | 1.67 | 1.10 | 1.24 |
| Average | 1.72 | 1.19 | 1.15 |

GRADE VI

| Test | | | |
|---|---|---|---|
| I | 1.24 | 1.17 | 1.04 |
| II | 1.36 | 1.23 | 1.83 |
| III | 1.24 | 1.04 | 1.13 |
| IV | 1.33 | .83 | .80 |
| V | 1.33 | .83 | 1.14 |
| VI | 1.57 | 1.56 | 1.60 |
| VII | 2.00 | 1.39 | 1.45 |
| VIII | 1.38 | 1.24 | 1.55 |
| IX | 1.20 | 1.32 | 1.34 |
| Com. Arith. | 1.35 | .94 | 1.66 |
| Average | 1.72 | 1.26 | 1.35 |

GRADE VII

| Test | | | |
|---|---|---|---|
| I | 1.18 | 1.06 | 1.05 |
| II | 1.43 | 1.38 | 1.55 |
| III | 1.20 | 1.05 | 1.33 |
| IV | 1.20 | 1.46 | 1.41 |
| V | 1.37 | 1.08 | 1.06 |
| VI | 1.19 | 1.41 | .94 |
| VII | 1.50 | 1.80 | 2.29 |
| VIII | 1.77 | 1.28 | 1.14 |
| IX | 1.27 | 1.28 | 1.21 |
| Com. Arith | 1.93 | 1.23 | |
| Average | 1.40 | 1.30 | 1.33 |

TABLE XV—Continued

GRADE VIII

| Test | October | February | May |
|---|---|---|---|
| I | 1.21 | 1.00 | 1.38 |
| II | 1.73 | 1.41 | 1.44 |
| III | 1.53 | 1.61 | 1.20 |
| IV | 1.18 | 1.06 | 1.04 |
| V | 1.23 | .95 | 1.00 |
| VI | 1.52 | 1.84 | 1.50 |
| VII | 1.73 | 1.27 | 1.66 |
| VIII | 1.72 | 1.17 | 1.31 |
| IX | 1.23 | 1.47 | 1.44 |
| Com. Arith. | 1.50 | 1.20 | 1.07 |
| Average | 1.46 | 1.30 | 1.00 |

The grade averages of the ratios of the quartiles are given in Table XIV. These also without exception decrease from October to February and from October to May, but from February to May, five of the seven averages show an increase. If we study the ratios for the tests for each grade which are given in Table XV we find that out of sixty sets 81 per cent show decrease from October to February and 76 per cent show decrease from October to May. We may, therefore, say that in over 75 per cent of the cases the quartile ratios show a decrease in variability during the course of training.

In Table XVI we find the grade averages of the ratios of the sixtieth percentile to the fortieth percentile scores. With one exception, all the averages indicate a decrease in variability from October to May, but from February to May, five of the seven averages show the general effect, namely, a decrease from 1.70 to 1.19 from October to February and a decrease from 1.70 to 1.16 from October to May. If we study the ratios from which the grade averages were calculated, as in Table XVI, we find that 70 per cent show decrease from October to May and 67 per cent show decrease from October to May. The same conclusion is therefore justified by this measure as from the others, but as it measures individuals who are closer together in the initial performance, the tendency toward decreased variability from practice is not so marked.

TABLE XVI

*Grade averages of ratios of 60th percentile to 40th percentile for October, February and May*

| Grade | October | February | May |
|---|---|---|---|
| II | 1.29 | 1.18 | 1.69 |
| III | 1.26 | 1.07 | 1.12 |
| IV | 1.28 | 1.21 | 1.23 |
| V | 1.23 | 1.84 | 1.07 |
| VI | 1.10 | 1.03 | 1.08 |
| VII | 1.09 | .97 | .89 |
| VIII | 1.15 | 1.02 | 1.05 |
| Final Average | 1.70 | 1.19 | 1.16 |

TABLE XVII

*Ratio of 60th percentile to 40th percentile for each test in each grade for October, February and May*

GRADE II

| | October | February | May |
|---|---|---|---|
| Total Reading | 1.25 | 1.48 | 1.41 |
| Total Arithmetic | 1.5 | 1.00 | .98 |
| Com. Arithmetic | 1.11 | 1.01 | 1.02 |
| Average | 1.29 | 1.18 | 1.69 |

GRADE III

| Test | | | |
|---|---|---|---|
| I | 1.25 | 1.48 | 1.41 |
| II | 1.44 | .99 | 1.54 |
| III | 1.75 | 1.21 | 1.12 |
| IV | 1.14 | 1.00 | 1.04 |
| V | 1.00 | 1.00 | 1.00 |
| VI | 1.21 | .91 | .85 |
| Com. Arith. | 1.05 | .88 | .92 |
| Average | 1.26 | 1.07 | 1.12 |

GRADE IV

| Test | | | |
|---|---|---|---|
| I | 1.15 | .96 | .96 |
| II | 1.42 | 1.20 | 1.88 |
| III | 1.44 | 1.28 | 1.07 |
| IV | 1.07 | 1.06 | 1.10 |
| V | 1.33 | 1.10 | 1.30 |
| VI | 1.27 | 1.73 | 1.93 |
| VII | 1.6 | 1.33 | 1.26 |
| VIII | 1.40 | 1.47 | .86 |
| IX | 1.07 | .94 | .89 |
| Com. Arith. | 1.05 | 1.11 | 1.11 |
| Average | 1.28 | 1.21 | 1.23 |

TABLE XVII—Continued

GRADE V

| Test | October | February | May |
|---|---|---|---|
| I | 1.22 | 1.24 | 1.24 |
| II | 1.20 | .71 | 1.29 |
| III | 1.17 | 1.14 | 1.36 |
| IV | 1.17 | .99 | .88 |
| V | 1.00 | 1.00 | 1.00 |
| VI | 1.27 | 1.03 | 1.24 |
| VII | 1.27 | .57 | .98 |
| VIII | 1.46 | .33 | .48 |
| IX | 1.17 | 1.42 | 1.34 |
| Com. Arith. | 1.33 | 1.00 | .90 |
| Average | 1.23 | 1.84 | 1.07 |

GRADE VI

| Test | | | |
|---|---|---|---|
| I | 1.03 | 1.11 | .90 |
| II | 1.05 | 1.07 | 1.52 |
| III | 1.02 | 1.27 | 1.56 |
| IV | 1.04 | .77 | .85 |
| V | 1.07 | .91 | 1.01 |
| VI | 1.04 | 1.21 | 1.13 |
| VII | 1.47 | .82 | .75 |
| VIII | 1.17 | .84 | .81 |
| IX | 1.11 | 1.31 | 1.26 |
| Com. Arith. | 1.00 | 1.00 | 1.00 |
| Average | 1.10 | 1.03 | 1.08 |

GRADE VII

| Test | | | |
|---|---|---|---|
| I | 1.08 | 1.16 | 1.03 |
| II | 1.06 | 1.26 | 1.04 |
| III | 1.04 | .85 | 1.14 |
| IV | 1.03 | .94 | .94 |
| V | 1.05 | .84 | 1.05 |
| VI | 1.08 | 1.23 | 1.18 |
| VII | 1.17 | .94 | .64 |
| VIII | 1.08 | .60 | .83 |
| IX | 1.09 | .97 | 1.03 |
| Com. Arith. | 1.26 | .88 | .... |
| Average | 1.09 | .97 | .89 |

GRADE VIII

| Test | | | |
|---|---|---|---|
| I | 1.08 | .95 | .97 |
| II | 1.13 | 1.24 | 1.34 |
| III | 1.10 | 1.02 | 1.08 |
| IV | 1.07 | 1.08 | 1.29 |
| V | 1.17 | .98 | .85 |
| VI | 1.17 | .98 | 1.13 |
| VII | 1.35 | .69 | .67 |
| VIII | 1.14 | 1.09 | 1.11 |
| IX | 1.11 | .98 | 1.00 |
| Com. Arith. | 1.13 | 1.17 | 1.07 |
| Average | 1.15 | 1.02 | 1.05 |

TABLE XVIII

*Ratios of average of three highest to average of three lowest for October and May scores, coefficients of variability for October and May scores, and correlations between October scores and per cent of gain from October to May.*

All measurements are by tests regardless of grades. Correlations from Edwards, 1929. S. D.'s and averages for calculation of coefficients of variability also supplied by him.

| | Ave. 3 Highest / Ave. 3 Lowest | | S. D. / Average | | Correlation October Score and |
|---|---|---|---|---|---|
| Test | Oct. | May | Oct. | May | Per cent Gain |
| I | 46.75 | 4.07 | .63 | .41 | —.60 |
| II | 32.40 | 5.33 | .68 | .55 | —.63 |
| III | 31.50 | 5.16 | .63 | .50 | —.66 |
| IV | 15.81 | 2.97 | .53 | .37 | —.75 |
| V | 11.93 | 4.59 | .68 | .47 | —.65 |
| VI | 40.15 | 4.93 | .66 | .34 | —.75 |
| VII | 22.03 | 4.45 | .69 | .55 | —.62 |
| VIII | 14.30 | 1.96 | .47 | .41 | —.62 |
| IX | 33.50 | 3.94 | .38 | .30 | —.42 |
| Compass Arith | 9.50 | 1.77 | .67 | .55 | —.59 |

In Table XVIII we find the results of three measurements calculated by tests regardless of grades. They are the ratios of the average of 3 highest to the average of the three lowest for the October and May scores, the coefficients of variability for the October and May scores, and the correlations between the October scores and the per cent of gain from October to May. In calculating the average of three lowest for the ratios only scores greater than zero were used. Each of these measures shows a marked reduction in variability from October to May for each test. The extremes of the distribution show a sharp convergence, the variability of the central two-thirds of the group is much reduced, and the correlations between the October scores and the per cent of gain are all negative. This remarkable consistency together with the results of the same measurements when calculated for the tests of each grade leaves no doubt that the effect of the training ordinarily given in public schools is to reduce the variability in the achievement of the pupils. The large reductions shown in the facts of Table XVIII indicates that the learning of the skills measured by these tests is not dependent upon the possession of a high intelligence or unusual native capacity and that the degree of achievement that one acquires with them is

largely a product of training—the opportunity to learn and the effectiveness of the study and of the instruction. It may be that this is true of most of the ways by which men and women secure a livelihood and happiness. If so, the responsibility of schools and of other agencies for training is very great indeed.

TABLE XIX

*Summary showing per cent of cases that indicate convergence or divergence for each of four measures.* From Edwards, 1929

| | October to February | | | October to May | | |
|---|---|---|---|---|---|---|
| | Converge No. | Diverge No. | Per cent of cases converging | Converge No. | Diverge No. | Per cent of cases converging |
| Ave. 3 highest / Ave. 3 lowest | 59 | 1 | 98 | 57 | 2 | 96 |
| S. D. / Mean | 49 | 15 | 76 | 54 | 9 | 85 |
| Q3 / Q1 | 49 | 11 | 81 | 45 | 14 | 76 |
| 60th percentile / 40th percentile | 42 | 18 | 70 | 40 | 19 | 67 |

In Table XIX we find a summary of all the measures used in this research except the correlations. The conclusions already stated are confirmed but one new point is brought into relief, namely, that the greater the difference between individuals in achievement is the greater the probability that equal training will reduce this difference. This is shown most clearly in the last column of Table XIX. The changes are 96 in 100 that equal practice will reduce the difference between the extremes of a group, 85 in 100 that it will reduce the variability of the central two-thirds, 76 in 100 that it will reduce the variability of the middle half, and 67 in 100 that it will reduce the variability of the central 40 per cent of the group. It is reasonable to suppose that a longer or more intensive training than was here investigated would increase these probabilities.

From his study of the tables presented above, the reader has probably noticed that all measures indicate that the variability

of the different groups was much more reduced from October to February than from February to May. If the conclusion drawn above is true it should follow that the reduction in variability is proportional to the amount of gain during the interval between tests, and if so, most of the gains made during the year must have been made during the first half year. An investigation of the original scores with this problem in mind shows that such was the case. The percentage of pupils below standard was reduced from 70 to 47 between October and February and from 70 to 41 between October and May. Further evidence in favor of this view is found by comparing the percentage of reduction in each grade of the coefficient of variability from October to May with the percentage of a standard year's gain in each grade as measured by the achievement tests. The figures for this comparison are given in Table XX.

TABLE XX

*Comparison of amount of reduction in variability with amount of gains by grades*

| | Grades | | | | | | |
|---|---|---|---|---|---|---|---|
| | II | III | IV | V | VI | VII | VIII |
| Percentage of decrease in variability..... | 63 | 47 | 22 | 24 | 00 | 33 | 12 |
| Percentage of standard year's gain..... | 190 | 125 | 124 | 100 | 68 | 136 | 93 |

While the parallelism between the two rows is not perfect it is so close as to indicate that the amount of reduction in variability is to a large extent a function of the amount of gain during the period of training. If this is correct, we may understand why the variability was more reduced during the first half year than during the second half year and why it was more reduced in some grades than in others. Incidentally the suggestion occurs to us that we might use reduction in variability in a grade as one measure of teaching success.

Coming now to the question which we started out to solve in this part of the paper, namely, what changes in variability in achievement in school subjects result from eight months training, as it is usually given in the public schools?, we can do no better than to quote Mr. Edwards' conclusions as stated in his master's thesis.

" We have measured the variability of various groups in reading, arithmetic, spelling, language usage, history, literature and nature study. All of these subjects are generally conceded to be the fundamental requirements for an education in our present civilization. Measuring the variability at the beginning, midpoint, and end of practice we find that equal practice makes individuals more alike in achievement although they differ in original ability. Our experiments show that individual differences grow less as practice continues.

" We realize that there are limitations to the tests which form the basis of our problems. The nearer a person approaches the upper limits of a test the possibilities for improvement become less. It may be that in traits where the possibilities for improvement are less limited, the results might be different.

" The tests were given under ordinary school conditions and no systematic efforts were made to give instruction according to ability or to have every pupil work to the limit of his capacity. It is possible that if every pupil had been encouraged to do a maximum amount of work that the results would show greater variability. We were more interested, however, in studying conditions as they exist in the average school.

" The P. E. could not be estimated with the measures which we used, except for the correlations and the coefficients of variability, and the reliability of the figures is uncertain but, since all of the measures agree, the conclusions seem pretty safely established.

" We do not advocate that the talented students should be deprived of any privileges or experiences which will aid them in their educational development but we do believe that the lower levels of society can be greatly helped by giving them the same opportunities for advancement that are now provided for the most favored classes. Our results show that the gospel of ' equal opportunities for all ' is well justified."

## SUMMARY

The problem of this paper was to find out the influence of training on changes in variability, or if equal training makes a group of individuals more alike or more different in their achievement.

The solution reached to this problem before 1923 by the experimental investigations of Thorndike, Wells, Donovan and Thorndike, Hahn and Thorndike, Chapman, Thurstone, Starch, Henmon, and Race may be fairly well summarized by the conclusion that Thorndike drew from his experiment published in 1916, namely: "The effect of equalizing opportunity is thus to increase individual differences. This result now found with many different functions, is perhaps the strongest argument that differences in achievement are largely due to differences in original capacity."

The papers published on the subject since 1922, particularly those by Gates, Reed, Kincaid, and Peterson and Barlow, draw an opposite conclusion, namely, that practice reduces individual differences.

This contradiction is due to the methods of measuring used by the experimenters. A critical analysis of these different possible measures shows that the following, when used together are reliable: ratio of highest to lowest at beginning of practice compared with ratio of same individuals at end of practice, standard deviation divided by the mean, and correlation between initial performance and relative gain. The following have *fair* reliability: Negative correlation between initial performance and gross gain, average percentage of gain by the highest fourth as compared with average of lowest fourth when the individuals in these sections are kept the same; the ratio of the average of highest fourth to average of lowest fourth, or of seventh octile to first octile, or of $Q_3$ to $Q_1$ at the beginning of practice compared with the ratio of the same individuals at the end of practice. The following measures have *little or no* validity and should not be made the basis of judgments of the effect of practice on indi-

vidual differences: correlations between initial and final performance, positive correlations between initial performance and gross gain; ratio of best to worst, next best to next worst, etc., at the beginning and at the end of practice; gross gains in amount of work per unit of time or amount of time per unit of work; and standard deviations and like measures when considered apart from their measures of central tendency.

The experimenters who drew the conclusion that practice increases individual differences based their judgments for the most part on invalid measures of variability. When their results are measured by the methods recommended in this paper, most of them show that practice reduces individual differences.

The measurement of gains by work per time unit or time per unit of work made in a new practice experiment in tossing and catching the yo-yo top shows that these measures lead to contradictory conclusions but that when the results are measured by ratio methods, the coefficient of variability, or the correlation between initial performance and relative gains, they consistently show a very sharp decrease in variability.

Variability in achievement in school subjects is reduced by the training ordinarily given in the public schools. This significant conclusion was drawn from an investigation made by C. J. Edwards of the training received by 137 school children from Grades II to VIII inclusive. Achievement was measured by the Compass Survey Arithmetic Tests and Stanford Achievement Tests which were given in October, February and May.

The farther apart two individuals are in achievement the greater the probability is that the difference between them will be reduced by equal practice.

The amount of reduction in variability is largely a function of the amount of gain during the period of practice.

Inequality in achievement in school subjects and in motor skill is reduced by giving pupils and students an equal opportunity to learn. If this is desirable, it justifies the gospel of equal opportunity as a policy in public education. It also appears that heredity has been over-emphasized as a factor in explaining individual differences in achievement.

The question of the effect on variability of training that is adjusted to the maximum capacity of each individual of a group is not considered in this investigation.

The conclusions stated above are dependent upon the measures of variability used in our researches. If they are invalid we hope that other investigators will find better ones and so discover more exact conclusions.

The above conclusions are also subject to modification by the solution of certain related problems. Among these are how we make the practise of a group of individuals the same in kind, in difficulty, and in amount; how we may measure achievement in practise on a scale which starts from an absolute zero and whose units in one part are comparable with the units in another; what the correct definition of variability is and what the correct measure of it is; and what the criteria are of a valid statistical measure. We hope that others interested in scientific research will contribute to the solution of these problems.

## BIBLIOGRAPHY

BARLOW, MYRON C. Individual differences as affected by continued practice. Master's Thesis, *George Peabody College for Teachers,* August, 1924.

BROWN, WARNER. Habit interference in sorting cards. *University of California Publications in Psychology,* 1914, I, No. 4, 269–321.

CHAPMAN, J. C. Individual differences in ability and improvement and their correlations. *Teachers College, Columbia University, Contributions to Education,* 1914, No. 63.

CHAPMAN, J. C. Learning curve in typewriting. *Journal of Applied Psychology,* 1919, III, 252–268.

CHAPMAN, J. C. Statistical considerations in interpreting the effect of training on individual differences. *Psychological Review,* 1925, 32, 224–234.

DONOVAN, M. E., and THORNDIKE, E. L. Improvement in a practice experiment under school conditions. *American Journal of Psychology,* 1913, 24, 426–428.

EDWARDS, C. J. The influence of training for changes in variability in achievement. Master's thesis. *Western State College,* Gunnison, Colorado.

GATES, G. S. Individual differences as affected by practice. *Archives of Psychology,* 1922, 8, No. 58.

HAHN, H. H., and THORNDIKE, E. L. Some results of practice in addition under school conditions. *Journal of Educational Psychology,* 1914, 5, 65–84.

HURD, A. W. Re-organization in physics. *North Central Association Quarterly,* 1929, 4, 277–293.

KINCAID, MARGARET. A study in individual differences in learning. *Psychological Review,* 1925, 33, 34–53.

McCall, W. A. Correlation of some psychological and educational measurements. *Teachers College, Columbia University, Contributions to Education,* 1916, No. 79.

Myers, G. C. Some variabilities and correlations in learning. *Journal of Educational Psychology,* 1918, 9, 316–326.

Perrin, F. A. C. The learning curves of the analogies and the mirror reading tests. *Psychological Review,* 1919, 27, 42–62.

Peterson, J. Experiments in ball tossing: The significance of learning curves. *Journal of Experimental Psychology,* 1917, 2, 178–224.

Peterson, J., and Barlow, M. C. The effects of practice on individual differences. *The Twenty-Seventh Yearbook of the National Society for the Study of Education,* Part II, 1928, 211–230.

Starch, Daniel. Educational Psychology, 1919.

Starch, Daniel. Experiments in educational psychology, 1911.

Race, Henrietta. Improvability: Its inter-correlations and its relation to initial ability. Teachers College, Columbia University, 1922.

Reed, H. B. The effect of training on individual differences. *Journal of Experimental Psychology,* 1924, 7, 186–201.

Stoddard, Geo. D. The problem of individual differences in learning. *Psychological Review,* 1925, 32, 479–485.

Thorndike, E. L. The effect of practice in case of a purely intellectual function. *American Journal of Psychology,* 1908, 19, 374–384.

Thorndike, E. L. Practice in case of addition. *American Journal of Psychology,* 1910, 21, 483–486.

Thorndike, E. L. Relation between initial ability and improvement in a substitution test. *School and Society,* 1915, I, 429–430.

Thorndike, E. L. Notes on practice, improvability and curve of work. *American Journal of Psychology,* 1916, 27, 550–565.

Thorndike, E. L. Educational Psychology, 1914, III.

Thurstone, L. L. Variability in learning. *Psychological Bulletin,* 1918, 15, 210–212. See criticism of this paper by J. Peterson on pp. 452–456.

Wells, F. L. The relation of practice to individual differences. *American Journal of Psychology,* 1912, 23, 75–88.

Whitley, M. T. An empirical study of certain tests for individual differences. *Columbia University, Archives of Psychology,* 1911, No. 19.

Woodrow, Herbert. Practice and transference in normal and feeble-minded children. *Journal of Educational Psychology,* 1917, 8, 85–96.